AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

S. HRG. 118–497

BIG TECH AND THE ONLINE CHILD SEXUAL EXPLOITATION CRISIS

HEARING

BEFORE THE

COMMITTEE ON THE JUDICIARY
UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

SECOND SESSION

JANUARY 31, 2024

Serial No. J–118–53

Printed for the use of the Committee on the Judiciary

BIG TECH AND THE ONLINE CHILD SEXUAL EXPLOITATION CRISIS

S. HRG. 118–497

BIG TECH AND THE ONLINE CHILD SEXUAL EXPLOITATION CRISIS

HEARING

BEFORE THE

COMMITTEE ON THE JUDICIARY
UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

SECOND SESSION

JANUARY 31, 2024

Serial No. J–118–53

Printed for the use of the Committee on the Judiciary

www.judiciary.senate.gov
www.govinfo.gov

U.S. GOVERNMENT PUBLISHING OFFICE
57–444 WASHINGTON : 2025

COMMITTEE ON THE JUDICIARY

RICHARD J. DURBIN, Illinois, *Chair*

SHELDON WHITEHOUSE, Rhode Island
AMY KLOBUCHAR, Minnesota
CHRISTOPHER A. COONS, Delaware
RICHARD BLUMENTHAL, Connecticut
MAZIE K. HIRONO, Hawaii
CORY A. BOOKER, New Jersey
ALEX PADILLA, California
JON OSSOFF, Georgia
PETER WELCH, Vermont
LAPHONZA BUTLER, California

LINDSEY O. GRAHAM, South Carolina, *Ranking Member*
CHARLES E. GRASSLEY, Iowa
JOHN CORNYN, Texas
MICHAEL S. LEE, Utah
TED CRUZ, Texas
JOSH HAWLEY, Missouri
TOM COTTON, Arkansas
JOHN KENNEDY, Louisiana
THOM TILLIS, North Carolina
MARSHA BLACKBURN, Tennessee

JOSEPH ZOGBY, *Majority Staff Director*
KATHERINE NIKAS, *Minority Staff Director*

CONTENTS

OPENING STATEMENTS

	Page
Durbin, Hon. Richard J.	1
Graham, Hon. Lindsey O.	3

WITNESSES

Chew, Shou	11
Prepared statement	69
Responses to written questions	102
Citron, Jason	6
Prepared statement	77
Responses to written questions	219
Spiegel, Evan	9
Prepared statement	87
Responses to written questions	314
Yaccarino, Linda	12
Prepared statement	91
Responses to written questions	393
Zuckerberg, Mark	7
Prepared statement	97
Responses to written questions	471

APPENDIX

Items submitted for the record	67

BIG TECH AND THE ONLINE CHILD SEXUAL EXPLOITATION CRISIS

WEDNESDAY, JANUARY 31, 2024

UNITED STATES SENATE,
COMMITTEE ON THE JUDICIARY,
Washington, DC.

The Committee met, pursuant to notice, at 10 a.m., in Room G50, Dirksen Senate Office Building, Hon. Richard J. Durbin, Chair of the Committee, presiding.

Present: Senators Durbin [presiding], Whitehouse, Klobuchar, Coons, Blumenthal, Hirono, Booker, Padilla, Ossoff, Welch, Butler, Graham, Cornyn, Lee, Cruz, Hawley, Cotton, Kennedy, Tillis, and Blackburn.

OPENING STATEMENT OF HON. RICHARD J. DURBIN, A U.S. SENATOR FROM THE STATE OF ILLINOIS

Chair DURBIN. This meeting of the Senate Judiciary Committee will come to order. I thank all those in attendance. I want to preface my remarks by saying that I've been in Congress for a few years. Senator Graham has as well. If you do not believe this is an idea whose time has come, take a look at the turnout here.

Today, the Senate Judiciary Committee will continue its work on an issue on the mind of most American families: how to keep our kids safe from sexual exploitation and harm in the internet age. Online child sexual exploitation includes the use of online platforms to target and groom children, and the production and endless distribution of child sexual abuse material, CSAM, which can haunt victims for their entire lives, and in some cases, take their lives.

Everyone here will agree this conduct is abhorrent. I'd like to turn to a brief video to hear directly from the victims, the survivors, about the impact these crimes have had on them.

[Video is shown.]

Chair DURBIN. Online child sexual exploitation is a crisis in America. In 2013, the National Center for Missing and Exploited Children, known as NCMEC, received approximately 1,380 cyber tips per day. By 2023, just 10 years later, the number of cyber tips has risen to 100,000 reports a day. That's a 100,000 daily reports of child sexual abuse material, also known as CSAM.

In recent years, we've also seen an explosion in the so-called financial sextortion, in which a predator uses a fake social media account to trick a minor into sending explicit photos or videos then threatens to release them unless the victim sends money.

In 2021, NCMEC received a total of 139 reports of sextortion. 2021. In 2023, through the end of October alone, this number skyrocketed to more than 22,000. More than a dozen children have died by suicide after becoming victims of this crime. This disturbing growth in child sexual exploitation is driven by one thing: changes in technology.

In 1996, the world's best-selling cell phone was the Motorola StarTAC. While groundbreaking at the time, the clamshell-style cell phone wasn't much different from a traditional phone. It allowed users to make and receive calls, and even receive text messages, but that was about it. Fast forward to today, smartphones are in the pockets of seemingly every man, woman, and teenager on the planet.

Like the StarTAC, today's smartphones allow users to make and receive calls and texts, but they can also take photos and videos, support live streaming, and offer countless apps. With the touch of your finger, that smartphone that can entertain and inform you can become a back alley where the lives of your children are damaged and destroyed. These apps have changed the ways we live, work, and play.

But as investigations have detailed, social media and messaging apps have also given predators powerful new tools to sexually exploit children. Your carefully crafted algorithms can be a more powerful force on the lives of our children than even the most best-intentioned parent.

Discord has been used to groom, abduct, and abuse children. Meta's Instagram helped connect and promote a network of pedophiles. Snapchat's disappearing messages have been co-opted by criminals who financially extort young victims. TikTok has become a "platform of choice for predators to access, engage, and groom children for abuse." And the prevalence of CSAM on X has grown as the company has gutted its trust and safety workforce.

Today, we'll hear from the CEOs of those companies. They're not only the tech companies that have contributed to this crisis, they're responsible for many of the dangers our children face online. Their design choices, their failures to adequately invest in trust and safety, their constant pursuit of engagement and profit over basic safety have all put our kids and grandkids at risk. Coincidentally, several of these companies implemented commonsense child safety improvements within the last week, days before their CEOs would have to justify their lack of action before this Committee.

But the tech industry alone is not to blame for the situation we're in. Those of us in Congress need to look in the mirror. In 1996, the same year the Motorola StarTAC was flying off shelves, and years before social media went mainstream, we passed Section 230 of the Communications Decency Act. This law immunized the then fledgling internet platforms from liability for user-generated content.

Interesting, only one other industry in America has an immunity from civil liability. We'll leave that for another day. For the past 30 years, Section 230 has remained largely unchanged, allowing Big Tech to grow into the most profitable industry in the history of capitalism without fear of liability for unsafe practices. That has to change.

Over the past year, this Committee has unanimously reported five bills that would finally hold tech companies accountable for child sexual exploitation on their platforms. Unanimous. Take a look at the composition and Membership of the Senate Judiciary Committee, and imagine if you will, there's anything we could agree on unanimously. These five bills were the objective of agreement. One of these bills is my STOP CSAM Act. Critically, it would let victims sue online providers that promote, or aid and abet online child sexual exploitation, or that host or store CSAM.

This stand against online child sexual exploitation is bipartisan and absolutely necessary. Let this hearing be a call to action that we need to get kids online safety legislation to the President's desk. I now turn to the Ranking Member, Senator Graham.

STATEMENT OF HON. LINDSEY O. GRAHAM,
A U.S. SENATOR FROM THE STATE OF SOUTH CAROLINA

Senator GRAHAM. Thank you, Mr. Chairman. The Republicans will answer the call. All of us. Every one of us is ready to work with you and our democratic colleagues on this Committee to prove to the American people that while Washington is certainly broken, there's a ray of hope, and it is here. It lies with your children.

After years of working on this issue with you and others, I've come to conclude the following: social media companies, as they're currently designed and operate, are dangerous products. They're destroying lives, threatening democracy itself. These companies must be reigned in or the worst is yet to come.

Brandon Guffey is a Representative—Republican Representative from South Carolina in the Rock Hill area. To all the victims who came and showed us photos of your loved ones, don't quit. It's working. You're making a difference. Through you, we'll get to where we need to go so other people won't have to show a photo of their family, the damage to your family's been done. Hopefully, we can take your pain and turn it into something positive so nobody else has to hold up a sign.

Brandon's son got online with the Instagram and was tricked by a group in Nigeria that put up a young lady posing to be his girlfriend. And as things go at that stage in life, he gave her some photos—compromising sexual photos—and it turned out that she was part of a extortion group in Nigeria. They threatened the young man that if you don't give us money, we're going to expose these photos.

He gave them money, but it wasn't enough. They kept threatening, and he killed himself. They threatened Mr. Guffey and his son. These are bastards by any known definition. Mr. Zuckerberg, you and the companies before us, I know you don't mean it to be so, but you have blood on your hands. You have a product——

[Applause.]

Senator GRAHAM. You have a product that's killing people. When we had cigarettes killing people, we did something about it. Maybe not enough. You are going to talk about guns, we have the ATF. Nothing here. There's not a damn thing anybody can do about it. You can't be sued.

Now, Senator Blumenthal and Blackburn, who've been like the dynamic duo here, have found emails from your company where

they warned you about this stuff, and you decided not to hire 45 people that could do a better job of policing this. So the bottom line is you can't be sued. You should be, and these emails would be great for punitive damages, but the courtroom's closed. Every American abused by all the companies in front of me. Of all the people in America we could give blanket liability protection too, this would be the last group I would pick.

[Applause.]

Senator GRAHAM. It is now time to repeal Section 230. This Committee is made up of ideologically the most different people you could find. We've come together through your leadership, Mr. Chairman, to pass five bills to deal with the problem of exploitation of children. I'll talk about them in depth in a little bit. The bottom line is, all these bills have met the same fate. They go nowhere. They leave the Committee and they die.

Now, there's another approach. What do you do with dangerous products? You either allow lawsuits, you have statutory protections to protect consumers, or you have a commission of sorts to regulate the industry in question; to take your license away if you have a license, to fine you.

None of that exists here. We live in America, in 2024, where there is no regulatory body dealing with the most profitable, biggest companies in the history of the world. They can't be sued, and there's not one law on the book that's meaningful protecting the American consumer. Other than that, we're in a good spot.

So here's what I think's going to happen. I think after this hearing today, we're going to put a lot of pressure on our colleagues' leadership of the Republican, Democratic Senate to let these bills get to the floor and vote. And I'm going to go down, starting in a couple of weeks, make unanimous consent request to do CSAM, do the EARN IT Act, do your bill, do all of the bills, and you can be famous. Come and object. I'm going to give you a chance to be famous.

Now, Elizabeth Warren and Lindsey Graham have almost nothing in common. I promised her I would say that publicly.

[Laughter.]

The only thing worse than me doing a bill with Elizabeth Warren is her doing a bill with me. We have sort of part that because Elizabeth and I see an abuse here that needs to be dealt with.

Senator Durbin and I have different political philosophies, but I appreciate what you've done on this Committee. You have been a great partner. To all of my Democratic colleagues, thank you very, very much.

[Applause.]

Senator GRAHAM. To my Republican colleagues, thank you all very, very much. Save the applause for when we get a result. This is all talk right now, but there will come a day if we keep pressing to get the right answer for the American people. What is that answer? Accountability.

Now, these products have an upside. You've enriched our lives in many ways. Mr. Zuckerberg, you created a product I use. The idea, I think, when you first came out of this, be able to talk to your friends and your family, and pass on your life to be able to have

a place where you could talk to your friends and family about good things going on in life. And I use it. We all use it.

There's an upside to everything here, but the dark side hasn't been dealt with. It's now time to deal with the dark side because people have taken your idea and they have turned it into a nightmare for the American people. They've turned it into a nightmare for the world at large.

TikTok, we had a great discussion about how maybe Larry Ellison through Oracle can protect American data from Chinese communist influence. But TikTok, your representative in Israel, quit the company because TikTok is being used in a way to basically destroy the Jewish state. This is not just about individuals. I worry that in 2024 our democracy will be attacked again through these platforms by foreign actors. We're exposed, and AI is just starting.

So to my colleagues, we're here for a reason. This Committee has a history of being tough, but also doing things that need to be done. This Committee has risen to the occasion. There's more that we can do, but to the Members of this Committee, let's insist that our colleagues rise to the occasion also. Let's make sure that in the 118th Congress, we have votes that would fix this problem. All you can do is cast your vote at the end of the day, but you can urge the system to require others to cast their vote.

Mr. Chairman, I will continue to work with you and everybody on this Committee to have a day of reckoning on the floor of the U.S. Senate. Thank you.

Chair DURBIN. Thank you, Senator Graham. Today, we welcome five witnesses whom I'll introduce now. Jason Citron, the CEO of Discord Incorporated. Mark Zuckerberg, the founder and CEO of Meta. Evan Spiegel, the co-founder and CEO of Snap Incorporated. Shou Chew, the CEO of TikTok, and Linda Yaccarino, the CEO of X Corporation, formerly known as Twitter.

I will note for the record that Mr. Zuckerberg and Mr. Chew are appearing voluntarily. I'm disappointed that our other witnesses did not offer that same degree of cooperation. Mr. Citron, Mr. Spiegel, and Ms. Yaccarino are here pursuant to subpoenas, and Mr. Citron only accepted services of his subpoena after U.S. Marshals were sent to Discord's headquarters at taxpayers' expense. I hope this is not a sign of your commitment or lack of commitment to addressing the serious issue before us.

After I swear in the witnesses, each witness will have 5 minutes to make an opening statement. Then, Senators will ask questions in an opening round each of 7 minutes. I expect to take a short break at some point during questioning to allow the witnesses to stretch their legs. If anyone is in need of a break at any point, please let my staff know.

Before I turn to the witnesses, I'd also like you to take a moment to acknowledge that this hearing has gathered a lot of attention, as we expected. We have a large audience, the largest I've seen in this room, today. I want to make clear, as with other Judiciary Committee hearings, we ask people to behave appropriately. I know there is high emotion in this room, for justifiable reasons, but I ask you to please follow the traditions of the Committee.

That means no standing, shouting, chanting, or applauding witnesses. Disruptions will not be tolerated. Anyone who does disrupt

the hearing will be asked to leave. The witnesses are here today to address a serious topic. We want to hear what they have to say. I thank you for your cooperation. Could all of the witnesses please stand to be sworn in?

[Witnesses are sworn in.]

Chair DURBIN. Let the record reflect that all the witnesses have answered in the affirmative. Mr. Citron, please proceed with your opening statement.

STATEMENT OF MR. JASON CITRON, CO-FOUNDER AND CHIEF EXECUTIVE OFFICER, DISCORD INCORPORATED, SAN FRANCISCO, CALIFORNIA

Mr. CITRON. Good morning.

Chair DURBIN. Good morning.

Mr. CITRON. My name is Jason Citron, and I am the co-founder and CEO of Discord. We are an American company with about 800 employees living and working in 33 States. Today, Discord has grown to more than 150 million monthly active users.

Discord is a communications platform where friends hang out and talk online about shared interests from fantasy sports to writing music to video games. I've been playing video games since I was 5 years old, and as a kid, it's how I had fun and found friendship. Many of my fondest memories are of playing video games with friends. We built Discord so that anyone could build friendships playing video games from Minecraft, to Wordle, and everything in between. Games have always brought us together, and Discord makes that happen today.

Discord is one of the many services that have revolutionized how we communicate with each other in the different moments of our lives; iMessage, Zoom, Gmail, and on and on. They enrich our lives; create communities; accelerate commerce, healthcare, and education.

Just like with all technology and tools, there are people who exploit and abuse our platforms for immoral and illegal purposes. All of us here on the panel today, and throughout the tech industry, have a solemn and urgent responsibility to ensure that everyone who uses our platforms is protected from these criminals, both online and off.

Discord has a special responsibility to do that because a lot of our users are young people. More than 60 percent of our active users are between the ages of 13 and 24. It's why safety is built into everything we do. It's essential to our mission and our business, and most of all, this is deeply personal. I'm a dad with two kids. I want Discord to be a product that they use and love, and I want them to be safe on Discord. I want them to be proud of me for helping to bring this product to the world.

That's why I'm pleased to be here today to discuss the important topic of the online safety of minors. My written testimony provides a comprehensive overview of our safety programs. Here are a few examples of how we protect and empower young people.

First, we've put our money into safety. The tech sector has a reputation of larger companies buying smaller ones to increase user numbers and boost financial results. But the largest acquisition we've ever made at Discord was a company called Sentropy. It

didn't help us expand our market share or improve our bottom line. In fact, because it uses AI to help us identify, ban, and report criminals and bad behavior, it has actually lowered our user count by getting rid of bad actors.

Second, you've heard of end-to-end encryption that blocks anyone, including the platform itself, from seeing users' communications. It's a feature on dozens of platforms but not on Discord. That's a choice we've made. We don't believe we can fulfill our safety obligations if the text messages of teens are fully encrypted because encryption would block our ability to investigate a serious situation, and when appropriate, report to law enforcement.

Third, we have a zero-tolerance policy on child sexual abuse material or CSAM. We scan images uploaded to Discord to detect and block the sharing of this abhorrent material. We've also built an innovative tool, Teen Safety Assist, that blocks explicit images and helps young people easily report unwelcome conversations. We've also developed a new semantic hashing technology for detecting novel forms of CSAM called Clip, and we're sharing this technology with other platforms through the tech coalition.

Finally, we recognize that improving online safety requires all of us to work together. So we partner with nonprofits, law enforcement, and our tech colleagues to stay ahead of the curve in protecting young people online. We want to be the platform that empowers our users to have better online experiences, to build true connections, genuine friendships, and to have fun.

Senators, I sincerely hope today is the beginning of an ongoing dialog that results in real improvements in online safety. I look forward to your questions and to helping the Committee learn more about Discord.

[The prepared statement of Mr. Citron appears as a submission for the record.]

Chair DURBIN. Thank you, Mr. Citron. Mr. Zuckerberg.

STATEMENT OF MR. MARK ZUCKERBERG,
FOUNDER AND CHIEF EXECUTIVE OFFICER,
META, MENLO PARK, CALIFORNIA

Mr. ZUCKERBERG. Chairman Durbin, Ranking Member Graham, and Members of the Committee, every day, teens and young people do amazing things on our services. These are apps to create new things, express themselves, explore the world around them, and feel more connected to the people they care about. Overall, teens tell us that this is a positive part of their lives, but some face challenges online, so we work hard to provide parents and teens support and controls to reduce potential harms.

Being a parent is one of the hardest jobs in the world. Technology gives us new ways to communicate with our kids and feel connected to their lives, but it can also make parenting more complicated, and it's important to me that our services are positive for everyone who uses them. We are on the side of parents everywhere working hard to raise their kids.

Over the last 8 years, we've built more than 30 different tools, resources, and features that parents can set time limits for their teens using our apps, see who they're following, or if they report someone for bullying. For teens, we've added nudges to remind

them when they've been using Instagram for a while, or if it's getting late and they should go to sleep, as well as ways to hide words or people without those people finding out. We put special restrictions on teen accounts on Instagram. By default, accounts for under 16s are set to private, have the most restrictive content settings, and can't be messaged by adults that they don't follow or people they aren't connected to.

With so much of our lives spent on mobile devices and social media, it's important to look into the effects on teen mental health and well-being. I take this very seriously. Mental health is a complex issue and the existing body of scientific work has not shown a cause or a link between using social media and young people having worse mental health outcomes.

A recent National Academies of Sciences report evaluated over 300 studies and found that research, "did not support the conclusion that social media causes changes in adolescent mental health at the population level." It also suggested that social media can provide significant positive benefits when young people use it to express themselves, explore and connect with others. Still, we're going to continue to monitor the research and use it to inform our roadmap.

Keeping young people safe online has been a challenge since the internet began, and as criminals evolve their tactics, we have to evolve our defenses too. We work closely with law enforcement to find bad actors and help bring them to justice, but the difficult reality is that no matter how much we invest or how effective our tools are, there are always more. There's always more to learn and more improvements to make, but we remain ready to work with Members of this Committee, industry, and parents to make the internet safer for everyone.

I'm proud of the work that our teams do to improve online child safety on our services and across the entire internet. We have around 40,000 people overall working on safety and security, and we've invested more than $20 billion in this since 2016, including around $5 billion in the last year alone. We have many teams dedicated to child safety and teen well-being, and we lead the industry in a lot of the areas that we're discussing today.

We built technology to tackle the worst online risks and share it to help our whole industry get better. Like Project Lantern, which helps companies share data about people who break child safety rules, and we're founding members of Take It Down, a platform which helps young people to prevent their nude images from being spread online.

We also go beyond legal requirements and use sophisticated technology to proactively discover abusive material, and as a result, we find and report more inappropriate content than anyone else in the industry. As the National Center for Missing and Exploited Children put it this week, "Meta goes above and beyond to make sure that there are no portions of their network where this type of activity occurs."

I hope we can have a substantive discussion today that drives improvements across the industry, including legislation that delivers what parents say they want—a clear system for age verification, and control over what apps their kids are using. Three

out of four parents want app store age verification, and four out of five want parental approval of whenever teens download apps. We support this. Parents should have the final say on what apps are appropriate for their children, and shouldn't have to upload their ID every time. That's what app stores are for.

We also support setting industry standards on age-appropriate content, and limiting signals for advertising to teens to of age and location and not behavior. At the end of the day, we want everyone who uses our services to have safe and positive experiences.

Before I wrap up, I want to recognize the families who are here today who have lost a loved one, or lived through some terrible things that no family should have to endure. These issues are important for every parent and every platform. I'm committed to continuing to work in these areas, and I hope we can make progress today.

[The prepared statement of Mr. Zuckerberg appears as a submission for the record.]

Chair DURBIN. Thank you. Mr. Spiegel.

STATEMENT OF MR. EVAN SPIEGEL,
CO-FOUNDER AND CHIEF EXECUTIVE OFFICER,
SNAP INCORPORATED, SANTA MONICA, CALIFORNIA

Mr. SPIEGEL. Chairman Durbin, Ranking Member Graham, and Members of the Committee, thank you for convening this hearing, and for moving forward important legislation to protect children online. I'm Evan Spiegel, the co-founder and CEO of Snap. We created Snapchat, an online service that is used by more than 800 million people worldwide to communicate with their friends and family.

I know that many of you have been working to protect children online since before Snapchat was created, and we are grateful for your long-term dedication to this cause, and your willingness to work together to help keep our community safe. I want to acknowledge the survivors of online harms and the families who are here today who have suffered the loss of a loved one. Words cannot begin to express the profound sorrow I feel that a service we designed to bring people happiness and joy has been abused to cause harm.

I want to be clear that we understand our responsibility to keep our community safe. I also want to recognize the many families who have worked to raise awareness on these issues, push for change, and collaborated with lawmakers on important legislation like the Cooper Davis Act, which can help save lives.

I started building Snapchat with my co-founder, Bobby Murphy, when I was 20 years old. We designed Snapchat to solve some of the problems that we experienced online when we were teenagers. We didn't have an alternative to social media. That meant pictures shared online were permanent, public, and subject to popularity metrics. It didn't feel very good.

We built Snapchat differently because we wanted a new way to communicate with our friends that was fast, fun, and private. A picture is worth a thousand words, so people communicate with images and videos on Snapchat. We don't have public likes or comments when you share your story with friends. Snapchat is private

by default, meaning that people need to opt in to add friends and choose who can contact them. When we built Snapchat, we chose to have the images and videos sent through our service delete by default.

Unlike prior generations who've enjoyed the privacy afforded by phone calls which aren't recorded, our generation has benefited from the ability to share moments through Snapchat that may not be picture perfect, but instead convey emotion without permanence. Even though Snapchat messages are deleted by default, we let everyone know that images and videos can be saved by the recipient.

When we take action on illegal or potentially harmful content, we also retain the evidence for an extended period, which allows us to support law enforcement and hold criminals accountable. To help prevent the spread of harmful content on Snapchat, we approve the content that is recommended on our service using a combination of automated processes and human review.

We apply our content rules consistently and fairly across all accounts. We run samples of our enforcement actions through quality assurance to verify that we're getting it right. We also proactively scan for known child sexual abuse material, drug-related content, and other types of harmful content, remove that content, deactivate and device block offending accounts, preserve the evidence for law enforcement and report certain content to the relevant authorities for further action.

Last year, we made 690,000 reports to the National Center for Missing and Exploited Children leading to more than 1,000 arrests. We also removed 2.2 million pieces of drug-related content, and blocked 705,000 associated accounts. Even with our strict privacy settings, content moderation efforts, proactive detection, and law enforcement collaboration, bad things can still happen when people use online services. That's why we believe that people under the age of 13 are not ready to communicate on Snapchat.

We strongly encourage parents to use the device-level parental controls on iPhone and Android. We use them in our own household, and my wife approves every app that our 13-year-old downloads. For parents who want more visibility and control, we built Family Center on Snapchat where you can view who your teen is talking to, review privacy settings, and set content limits. We have worked for years with Members of the Committee on legislation like the Kids Online Safety Act and the Cooper Davis Act, which we are proud to support.

I want to encourage broader industry support for legislation protecting children online. No legislation is perfect, but some rules of the road are better than none. Much of the work that we do to protect people that use our service would not be possible without the support of our partners across the industry, government, nonprofit organizations, NGO's, and in particular, law enforcement and the first responders who have committed their lives to helping keep people safe.

I'm profoundly grateful for the extraordinary efforts across our country and around the world to prevent criminals from using online services to perpetrate their crimes. I feel an overwhelming sense of gratitude for the opportunities that this country has afforded me and my family. I feel a deep obligation to give back and

to make a positive difference, and I'm grateful to be here today as part of this vitally important democratic process.

Members of the Committee, I give you my commitment that we'll be part of the solution for online safety. We'll be honest about our shortcomings, and we'll work continuously to improve. Thank you, and I look forward to answering your questions.

[The prepared statement of Mr. Spiegel appears as a submission for the record.]

Chair DURBIN. Thank you, Mr. Spiegel. Mr. Chew.

STATEMENT OF MR. SHOU CHEW, CHIEF EXECUTIVE OFFICER, TIKTOK INCORPORATED, SINGAPORE

Mr. CHEW. Chair Durbin, Ranking Member Graham, and Members of the Committee, I appreciate the opportunity to appear before you today. My name is Shou Chew, and I'm the CEO of TikTok, an online community of more than 1 billion people worldwide, including well over 170 million Americans who use our app every month to create, to share, and to discover.

Now, although the average age on TikTok in the U.S. is over 30, we recognize that special safeguards are required to protect minors, and especially, when it comes to combating all forms of CSAM. As a father of three young children myself, I know that the issues that we're discussing today are horrific and the nightmare of every parent. I am proud of our efforts to address the threats to young people online from a commitment to protecting them, to our industry leading policies, use of innovative technology, and significant ongoing investments in trust and safety to achieve this goal.

TikTok is vigilant about enforcing its 13-and-up age policy, and offers an experience for teens that is much more restrictive than you and I would have as adults. We make careful product design choices to help make our app inhospitable to those seeking to harm teens. Let me give you a few examples of long-standing policies that you need to TikTok. We didn't do them last week.

First, direct messaging is not available to any users under the age of 16. Second, accounts for people under 16 are automatically set to private along with their content. Furthermore, the content cannot be downloaded and will not be recommended to people they do not know. Third, every teen under 18, has a screen time limit automatically set to 60 minutes. And fourth, only people 18 and above are allowed to use our livestream feature.

I'm proud to say that TikTok was among the first to empower parents to supervise their teens on our app with our family pairing tools. This includes setting screen time limits, filtering out content from the teens' feeds, amongst others. We made these choices after consulting with doctors and safety experts who understand the unique stages of teenage development to ensure that we have the appropriate safeguards to prevent harm and minimize risk.

Now, safety is one of the core priorities that defines TikTok under my leadership. We currently have more than 40,000 trust and safety professionals working to protect our community globally, and we expect to invest more than $2 billion in trust and safety efforts this year alone, with a significant part of that in our U.S. operations. Our robust community guidelines strictly prohibit con-

tent or behavior that puts teenagers at risk of exploitation or other harm, and we vigorously enforce them.

Our technology moderates all content uploaded to our app to help quickly identify potential CSAM and other material that breaks our rules. It automatically removes the content or elevates it to our safety professionals for further review. We also moderate direct messages for CSAM and related material, and use third-party tools like photo DNA and take it down to combat CSAM to prevent content from being uploaded to our platform.

We continually meet with parents, teachers, and teens. In fact, I sat down with a group just a few days ago. We use their insight to strengthen the protections on our platform, and we also work with leading groups like the Technology Coalition.

The steps that we're taking to protect teens are a critical part of our larger trust and safety work as we continue our voluntary and unprecedented efforts to build a safe and secure data environment for U.S. users, ensuring that our platform remains free from outside manipulation and implementing safeguards on our content recommendation and moderation tools.

Keeping teens safe online requires a collaborative effort as well as collective action. We share the Committee's concern and commitment to protect young people online, and we welcome the opportunity to work with you on legislation to achieve this goal. Our commitment is ongoing and unwavering because there is no finish line when it comes to protecting teens.

Thank you for your time and consideration today. I'm happy to answer your questions.

[The prepared statement of Mr. Chew appears as a submission for the record.]

Chair DURBIN. Thanks, Mr. Chew. Ms. Yaccarino.

STATEMENT OF MS. LINDA YACCARINO, CHIEF EXECUTIVE OFFICER, X CORP., SAN FRANCISCO, CALIFORNIA

Ms. YACCARINO. Chairman Durbin, Ranking Member Graham, and esteemed Members of the Committee, thank you for the opportunity to discuss X's work to protect the safety of minors online.

Today's hearing is titled a crisis which calls for immediate action. As a mother, this is personal, and I share the sense of urgency. X is an entirely new company, an indispensable platform for the world and for democracy. You have my personal commitment that X will be active and a part of this solution.

While I joined X only in June 2023, I bring a history of working together with governments, advocates, and NGO's to harness the power of media to protect people. Before I joined, I was struck by the leadership steps this new company was taking to protect children. X is not the platform of choice for children and teens.

We do not have a line of business dedicated to children. Children under the age of 13 are not allowed to open an account. Less than 1 percent of the U.S. users on X are between the ages of 13 and 17, and those users are automatically set to a private default setting, and cannot accept a message from anyone they do not approve.

In the last 14 months, X has made material changes to protect minors. Our policy is clear, X has zero tolerance toward any mate-

rial that features or promotes child sexual exploitation. My written testimony details X's extensive policies on content or actions that are prohibited, and include grooming, blackmail, and identifying alleged victims of CSE.

We've also strengthened our enforcement with more tools and technology to prevent those bad actors from distributing, searching for, and engaging with CSE content. If CSE content is posted on X, we remove it, and now we also remove any account that engages with CSE content, whether it is real or computer generated.

Last year, X suspended 12.4 million accounts for violating our CSE policies. This is up from 2.3 million accounts that were removed by Twitter in 2022. In 2023, 850,000 reports were sent to NCMEC, including our first ever autogenerated report. This is eight times more than was reported by Twitter in 2022.

We've changed our priorities. We've restructured our trust and safety teams to remain strong and agile. We are building a trust and safety center of excellence in Austin, Texas to bring more agents in-house to accelerate our impact. We're applying to the Technology Coalition's project, Lantern, to make further industry-wide progress and impact. We've also opened up our algorithms for increased transparency. We want America to lead in this solution.

X commends the Senate for passing the REPORT Act, and we support the SHIELD Act. It is time for a Federal standard to criminalize the sharing of nonconsensual intimate material. We need to raise the standards across the entire internet ecosystem, especially for those tech companies that are not here today and not stepping up. X supports the STOP CSAM Act. The Kids Online Safety Act should continue to progress, and we will support the continuation to engage with it and ensure the protections of the freedom of speech.

There are two additional areas that require everyone's attention. First, as the daughter of a police officer, law enforcement must have the critical resources to bring these bad offenders to justice. Second, with artificial intelligence, offenders' tactics will continue to sophisticate and evolve. Industry collaboration is imperative here.

X believes that the freedom of speech and platform safety can and must coexist. We agree that now is the time to act with urgency. Thank you. I look forward to answering your questions.

[The prepared statement of Ms. Yaccarino appears as a submission for the record.]

Chair DURBIN. Thank you very much, Ms. Yaccarino. Now we'll go into rounds of questions. Seven minutes each for the Members as well. I would like to make note of your testimony, Ms. Yaccarino, I believe you are the first social media company to publicly endorse the CSAM Act.

Ms. YACCARINO. It is our honor, Chairman.

Chair DURBIN. That is progress, my friends. Thank you for doing that. I'm still going to be asking some probing questions, but let me get down to the bottom line here. I'm going to focus on my legislation on CSAM. What it says is civil liability if you intentionally or knowingly host or store child sexual abuse materials or make child sex abuse materials available. Second, intentionally or knowingly promote, or aid and abet a violation of child sexual exploi-

tation laws. Is there anyone here who believes you should not be held civilly liable for that type of conduct? Mr. Citron.

Mr. CITRON. Good morning, Chair. You know, we very much believe that this content is disgusting and that there are many things about the STOP CSAM bill that I think are very encouraging and we very much support adding more resources for the CyberTipline and modernizing that along with giving more resources to NCMEC. And I'd be very open to having conversations with you and your team to talk through the details of the bills and more.

Chair DURBIN. I sure would like to do that because if you intentionally or knowingly host or store CSAM, I think you ought to at least be civilly liable. I can't imagine anyone who would disagree with that.

Mr. CITRON. Yes, it's disgusting content.

Chair DURBIN. It certainly is. That's why we need you supporting this legislation. Mr. Spiegel, I want to tell you, I listened closely to your testimony here, and it's never been a secret that Snapchat is used to send sexually explicit images. In 2013, early in your company's history, you admitted this in an interview. Do you remember that interview?

Mr. SPIEGEL. Senator, I don't recall this specific interview.

Chair DURBIN. You said that when you were first trying to get people on the app, you would, "go up to the people and be, like, hey, you should try this application. You can send disappearing photos. And they would say, oh, for sexting." Do you remember that interview?

Mr. SPIEGEL. Senator, when we first created the application, it was actually called Peekaboo, and the idea was around disappearing images. The feedback we received from people using the app is that they were actually using it to communicate. So we changed the name of the application to Snapchat, and we found that people were using it to talk visually.

Chair DURBIN. As early as 2017, law enforcement identified Snapchat as the pedophiles go-to sexual exploitation tool. The case of a 12-year-old girl identified in court only as LW shows the danger. Over 2½ years, a predator sexually groomed her, sending her sexually explicit images and videos over Snapchat.

The man admitted that he only used Snapchat with LW and not any other platforms because he, "knew the chats would go away." Did you and everyone else at Snap really fail to see that the platform was the perfect tool for sexual predators?

Mr. SPIEGEL. Senator, that behavior is disgusting and reprehensible. We provide in-app reporting tools so that people who are being harassed or who, you know, have been shared inappropriate sexual content can report it in the case of harassment or sexual content. We typically respond to those reports within 15 minutes so that we can provide help.

Chair DURBIN. When LW, the victim, sued Snapchat, her case was dismissed under Section 230 of the Communications Decency Act. Do you have any doubt that had Snap faced the prospect of civil liability for facilitating sexual exploitation, the company would've implemented even better safeguards?

Mr. SPIEGEL. Senator, we already work extensively to proactively detect this type of behavior. We make it very difficult for predators

to find teens on Snapchat. There are no public friends lists, no public profile photos. When we recommend friends for teens, we make sure that they have several mutual friends in common before making that recommendation. We believe those safeguards are important to preventing predators from misusing our platform.

Chair DURBIN. Mr. Citron, according to Discord's website, it takes, "a proactive and automated approach to safety only on servers with more than 200 members. Smaller servers rely on server owners and community moderators to define and enforce behavior."

So how do you defend an approach to safety that relies on groups of fewer than 200 sexual predators to report themselves for things like grooming, trading in CSAM, or sextortion?

Mr. CITRON. Chair, our goal is to get all of that content off of our platform, and ideally prevent it from showing up in the first place, or from people engaging in these kinds of horrific activities. We deploy a wide array of techniques that work across every surface on Discord.

I mentioned we recently launched something called Teen Safety Assist, which works everywhere, and it's on by default for teen users that kind of acts like a buddy that lets them know if they're in a situation or talking with someone that may be inappropriate so they can report that to us and block that user. So we——

Chair DURBIN. Mr. Citron, if that were working, we wouldn't be here today.

Mr. CITRON. Chair, this is an ongoing challenge for all of us. That that is why we're here today. But we do have—15 percent of our company is focused on trust and safety, of which this is one of our top issues. That's more people than we have working on marketing and promoting the company. So we take these issues very seriously, but we know it's an ongoing challenge, and I look forward to working with you and collaborating with our tech peers and the nonprofits to improve our approach.

Chair DURBIN. I certainly hope so. Mr. Chew, your organization, business is one of the more popular ones among children. Can you explain to us what you are doing particularly, and whether you've seen any evidence of CSAM in your business?

Mr. CHEW. Yes, Senator. We have a strong commitment to invest in trust and safety. And as I said in my opening statement, I intend to invest more than $2 billion in trust and safety this year alone. We have 40,000 safety professionals, you know, working on this topic. We have built a specialized child safety team to help us identify specialized issues, horrific issues, like material like the ones you have mentioned. If we identify any on our platform and we proactively do detection, we will remove it, and we will report them to NCMEC and other authorities.

Chair DURBIN. Why is it TikTok allowing children to be exploited into performing commercialized sex acts?

Mr. CHEW. Senator, I respectfully disagree with that characterization. Our live streaming product is not for anyone below the age of 18. We have taken action to identify anyone who violates that, and we remove them from using that service.

Chair DURBIN. At this point, I'm going to turn to my Ranking Member, Senator Graham.

Senator GRAHAM. Thank you, Mr. Chairman. Mr. Citron, you said we need to start a discussion. To be honest with you, we've been having this discussion for a very long time. We need to get a result, not a discussion. Do you agree with that?

Mr. CITRON. Ranking Member, I agree this is an issue that we've also been very focused on since we started our company in 2015, but this is first time we——

Senator GRAHAM. Are you familiar with the EARN IT Act, authored by myself and Senator Blumenthal?

Mr. CITRON. A little bit. Yes.

Senator GRAHAM. Okay. Do you support that?

Mr. CITRON. We——

Senator GRAHAM. Like, yes or no.

Mr. CITRON. We're not prepared to support it today, but we believe that Section——

Senator GRAHAM. Do you support the CSAM Act?

Mr. CITRON. The STOP CSAM Act, we are not prepared to support it today either.

Senator GRAHAM. Do you support the SHIELD Act?

Mr. CITRON. We believe that the CyberTipline——

Senator GRAHAM. Do you support it? Yes, or no?

Mr. CITRON. We believe that the CyberTipline and NCMEC——

Senator GRAHAM. I'll take that to be no. The Project Safe Childhood Act. Do you support it?

Mr. CITRON. We believe that——

Senator GRAHAM. I'll take that to be no. The REPORT Act. Do you support it?

Mr. CITRON. Ranking Member Graham, we very much look forward to having conversations with you and your team——

Senator GRAHAM. We look forward to passing the bill that will solve the problem. Do you support removing Section 230 liability protections for social media companies?

Mr. CITRON. I believe that Section 230 needs to be updated. It's a very old law.

Senator GRAHAM. Do you support repealing it so people can sue if they believe they're harmed?

Mr. CITRON. I think that Section 230 as written, while it has many downsides, has enabled innovation on the internet, which I think has largely been——

Senator GRAHAM. Thank you very much. So here you are. You got—if you're waiting on these guys to solve the problem, we're going to die waiting. Mr. Zuckerberg, I'll try to be respectful here. The Representative from South Carolina, Mr. Guffey's son, got caught up in a sex extortion ring in Nigeria using Instagram. He was shaken down, paid money that wasn't enough, and he killed himself using Instagram. What would you like to say to him?

Mr. ZUCKERBERG. It's terrible. I mean, no one should have to go through something like that.

Senator GRAHAM. You think he should be allowed to sue you?

Mr. ZUCKERBERG. I think that they can sue us.

Senator GRAHAM. Well, I think you should, and he can't. So the bottom line here, folks, is that this Committee is done with talking. We passed five bills unanimously that in their different ways—and look at who did this. Senators Graham and Blumenthal, Senators

Durbin and Hawley, Senators Klobuchar and Cornyn, Senators Cornyn and Klobuchar, and Senators Blackburn and Ossoff. I mean, we've found common ground here that just is astonishing. And we've had hearing after hearing, Mr. Chairman. And the bottom line is, I've come to conclude gentlemen, that you're not going to support any of this. Linda, how do you say your last name?

Ms. YACCARINO. Yaccarino.

Senator GRAHAM. Do you support the EARN IT Act?

Ms. YACCARINO. We strongly support the collaboration to raise industry——

Senator GRAHAM. No, no——

Ms. YACCARINO [continuing]. Practices to prevent CSAM.

Senator GRAHAM [continuing]. No, no. Do you support the EARN IT Act? In English, do you support the EARN IT Act? Yes, or no? We don't need double speak here.

Ms. YACCARINO. We look forward to supporting and continue our conversations. As you can see——

Senator GRAHAM. Okay. So I take that as no. But you have taken—the reason the EARN IT Act's important, you can actually lose your liability protection when children are exploited and you didn't use best business practices. See, the EARN IT Act means you have to earn liability protection. You aren't given it no matter what you do.

So to the Members of this Committee, it is now time to make sure that the people who are holding up the signs can sue on behalf of their loved ones. Nothing will change until the courtroom door is open to victims of social media. $2 billion, Mr. Chew, what percentage is that of what you made last year?

Mr. CHEW. Senator, it's a significant and increasing investment. As a private company, we're not——

Senator GRAHAM. You pay taxes. I mean, 2 percent is what percent of your revenue?

Mr. CHEW [continuing]. Senator, we're not ready to share our financials in public.

Senator GRAHAM. Well, I just think $2 billion sounds a lot unless you make a $100 billion. So the point is, you know, when you tell us you're going to spend $2 billion, great, but how much do you make? You know, it's all about eyeballs. Well, our goal is to get eyeballs on you.

And it's just not about children, I mean, the damage being done. Do you realize, Mr. Chew, that your TikTok representative in Israel resigned yesterday?

Mr. CHEW. Yes, I'm aware.

Senator GRAHAM. Okay. And he said, "I resigned from TikTok. We're living at a time in which our existence as Jews in Israel, and Israel is under attack and in danger. Multiple screenshots taken from TikTok's internal employee chat platform known as Lark, show how TikTok's trust and safety officers celebrate the barbaric acts of Hamas and other Iranian-back terror groups, including Houthis in Yemen."

Mr. CHEW. Senator, I need to make it very clear that pro-Hamas content and hate speech is not allowed at all——

Senator GRAHAM. Why did——

Mr. CHEW [continuing]. In our company.

Senator GRAHAM [continuing]. He resign—why did he resign? Why did he quit?

Mr. CHEW. Senator, we also do not allow any people——

Senator GRAHAM. Do you why did he quit?

Mr. CHEW. We do not allow this. We will investigate such claims——

Senator GRAHAM. But my question is, he quit. I'm sure he had a good job. He gave up a good job because he thinks your platform is being used to help people who want to destroy the Jewish state, and I'm not saying you want that. Mr. Zuckerberg, I'm not saying you want, as an individual, any of the harms. I am saying that the products you have created with all the upside have a dark side.

Mr. Citron, I am tired of talking. I'm tired of having discussions. We all know the answer here, and here's the ultimate answer: stand behind your product. Go to the American courtroom and defend your practices. Open up the courthouse door. Until you do that, nothing will change.

Until these people can be sued for the damage they're doing, it is all talk. I'm a Republican who believes in free enterprise, but also believe that every American who's been wronged has to have somebody to go to to complain. There's no commission to go to that can punish you. There's not one law in the book because you oppose everything we do and you can't be sued. That has to stop, folks.

How do you expect the people in the audience to believe that we're going to help their families if we don't have some system or a combination of systems to hold these people accountable? Because for all the upside, the dark side is too great to live with. We do not need to live this way as Americans.

Chair DURBIN. Thank you, Senator Graham. Senator Klobuchar is next. She's been quite a leader on the subject for quite a long time on the SHIELD Act and with Senator Cornyn on the revenge porn legislation. Senator Klobuchar.

Senator KLOBUCHAR. Thank you very much, Chairman Durbin, and thank you Ranking Member Graham for those words. I couldn't agree more. For too long we have been seeing the social media companies turn a blind eye when kids have joined these platforms in record numbers.

They have used algorithms that push harmful content because that content got popular. They provided a venue, maybe not knowingly at first, but for dealers to sell deadly drugs like fentanyl. Our own head of our Drug Enforcement Administration has said they basically have been captured by the cartels in Mexico and in China.

So I strongly support, first of all, the STOP CSAM bill. I agree with Senator Graham that nothing is going to change unless we open up the courtroom doors. I think the time for all of this immunity is done, because I think money talks even stronger than we talk up here.

Two of the five bills, as noted, are my bills with Senator Cornyn. One has actually passed through the Senate, but is waiting action in the House. But the other one is the SHIELD Act, and I do support appreciate those supportive of that bill. This is about revenge porn. The FBI Director testified before this Committee, there has been over 20 suicides of kids attributed to online revenge porn in just the last year.

But for those parents out there and those families, this is for them, about their own child, but it's also about making sure this doesn't happen to other children. I know because I've talked to these parents. Parents like Bridget Norring from Hastings, Minnesota, who is out there today. Bridget lost her teenage son after he took a fentanyl-laced pill that he purchased on the internet. Amy Neville is also here. Platform, got the pill. Amy Neville is also here. Her son, Alexander, was only 14 when he died after taking a pill he didn't know was actually fentanyl.

We're starting a law enforcement campaign, "One pill kills," in Minnesota, going to the schools with the sheriffs and law enforcement. But the way to stop it is, yes, at the border and at the points of entry, but we know that 30 percent, some of the people that are getting the fentanyl are getting it off the platforms.

Meanwhile, social media platforms generated $11 billion in revenue in 2022 from advertising directed at children and teenagers, including nearly $2 billion in ad profits derived from users age 12 and under. When a Boeing plane lost a door in mid-flight several weeks ago, nobody questioned the decision to ground a fleet of over 700 planes. So why aren't we taking the same type of decisive action on the danger of these platforms when we know these kids are dying?

We have bills——

[Applause.]

Senator KLOBUCHAR [continuing]. That have passed through this incredibly diverse Committee when it comes to our political views that have passed through this Committee, and they should go to the floor. We should do something finally about liability, and then we should turn to some of the other issues that a number of us have worked on when it comes to the charges for app stores, and when it comes to some of the monopoly behavior and the self-preferencing. But I'm going to stick with this today.

Facts: one-third of fentanyl cases investigated over 5 months, had direct ties to social media. That's from the DEA. Facts: between 2012 and 2022, CyberTipline reports of online child sexual exploitation increased from 415,000 to more than 32 million. And as I noted, at least 20 victims committed suicide in sextortion cases.

So, I'm going to start with that with you, Mr. Citron. My bill with Senator Cornyn, the SHIELD Act, includes a threat provision that would help protection and accountability for those that are threatened by these predators. Young kids get a picture, send it in, think they got a new girlfriend, or a new boyfriend, ruins their life or they think it's going to be ruined, and they kill themselves. So could you tell me why you're not supporting the SHIELD Act?

Mr. CITRON. Senator, we think it's very important that teens have a safe experience on our platforms. I think that the portion to strengthen law enforcement's ability to investigate crimes against children and hold bad actors accountable is incredible.

Senator KLOBUCHAR. So are you holding open that you may support it?

Mr. CITRON. We very much would like to have conversations with you. We're open to discussing further, and we do welcome legisla-

tion regulation. You know, this is a very important issue for our country, and you know, we've been prioritizing safety for——

Senator KLOBUCHAR. Okay, thank you.

Mr. CITRON [continuing]. Teens——

Senator KLOBUCHAR. I'm much more interested in if you support it because there's been so much talk at these hearings, and popcorn throwing, and the like, and I just want to get this stuff done. I'm so tired of this. It's been 28 years, what, since the internet—we haven't passed any of these bills because everyone's double-talk, double talk. It's time to actually pass them. And the reason they haven't passed is because of the power of your company. So let's be really, really clear about that. So what you say matters. Your words matter.

Mr. Chew, I'm a co-sponsor of Chair Durbin's STOP CSAM Act of 2023, along with Senator Hawley, who's the lead Republican, I believe, which, among other things, empowers victims by making it easier for them to ask tech companies to remove the material and related imagery from their platforms. Why would you not support this bill?

Mr. CHEW. Senator, we largely support it. I think the spirit of it is very aligned with what we want to do. There are questions about implementation that I think companies like us and some other groups have, and we look forward to asking those. And of course, if this legislation is law, we will comply.

Senator KLOBUCHAR. Mr. Spiegel, I know we talked ahead of time. I do appreciate your company's support for the Cooper Davis Act which will finally—it's a bill with Senator Shaheen and Marshall, which will allow law enforcement to do more when it comes to fentanyl. I think you know what a problem this is. Devin Norring, a teenager from Hastings—I mentioned his mom is here—suffered dental pain and migraine. So he bought what he thought was a percocet over Snap, but instead he bought a counterfeit drug laced with a lethal dose of fentanyl.

As his mom, who is here with us today said, "All of the hopes and dreams we as parents had for Devin were erased in the blink of an eye, and no mom should have to bury their kid." Talk about why you support the Cooper Davis Act.

Mr. SPIEGEL. Senator, thank you. We strongly support the Cooper Davis Act, and we will believe it will help DEA go after the cartels, and get more dealers off the streets to save more lives.

Senator KLOBUCHAR. Okay. Are there others that support that bill on this? No? Okay. Last, Mr. Zuckerberg. In 2021, The Wall Street Journal reported on internal Meta research documents asking, "Why do we care about tweens?" These were internal documents. I'm quoting the documents. And answering its own question by citing Meta internal emails, "They are a valuable but untapped audience."

At a commerce hearing, I'm also on that Committee, I asked Meta's head of global safety why children age 10 to 12 are so valuable to Meta. She responded, "We do not knowingly attempt to recruit people who aren't old enough to use our apps." Well, when the 42 State attorneys general, Democrat and Republican, brought their case they said this statement was inaccurate.

Few examples. In 2021, she received an email—Ms. Davis—from Instagram's research director saying that Instagram is investing in experiencing targeting young age, roughly 10 to 12. In a February 2021 instant message, one of your employees wrote that Meta is working to recruit Gen Alpha before they reach teenage years. A 2018 email that circulated inside Meta says that you were briefed that children under 13 will be critical for increasing the rate of acquisition when users turn 13.

Explain that, with what I heard at that testimony at the commerce hearing, that they weren't being targeted. And I just ask, again, as the other witnesses were asked, why your company does not support the STOP CSAM Act or the SHIELD Act?

Mr. ZUCKERBERG. Sure, Senator, I'm happy to talk to—to both of those. We had discussions internally about whether we should build a kids' version of Instagram, like the kids' version——

Senator KLOBUCHAR. I remember that.

Mr. ZUCKERBERG [continuing]. Of YouTube and other services. We haven't actually moved forward with that, and we currently have no plans to do so. So I can't speak directly to the exact emails that you cited, but it sounds to me like they were deliberations around a project that people internally thought was important and we didn't end up moving forward with.

Senator KLOBUCHAR. Okay. And the bills.

Mr. ZUCKERBERG. Yes.

Senator KLOBUCHAR. What are you going to say about the two bills?

Mr. ZUCKERBERG. Sure. So overall, I mean, my position on the bills is I agree with the goal of all of them. There are most things that I agree with within them. There are specific things that I would probably do differently. We also have our own legislative proposal for what we think would be most effective in terms of helping the internet and the various companies give parents control over the experience. So I'm happy to go into the detail on any one of them, but ultimately, I mean, I think that this is——

Senator KLOBUCHAR. Well, I think these parents will tell you that this stuff hasn't worked, to just give parents control. They don't know what to do. It's very, very hard, and that's why we are coming up with other solutions that we think are much more helpful to law enforcement, but also this idea of finally getting something going on liability. Because I just believe with all the resources you have, that you actually would be able to do more than you're doing, or these parents wouldn't be sitting behind you right now in this Senate hearing room.

Chair DURBIN. Thank you——

Mr. ZUCKERBERG. Senator Klobuchar——

Chair DURBIN [continuing]. Senator Klobuchar.

Mr. ZUCKERBERG [continuing]. Could I speak to that, or do you want me to come back later?

Chair DURBIN. Please, go ahead.

Mr. ZUCKERBERG. I don't think that parents should have to upload an ID or prove that they're the parent of a child in every single app that their children use. I think the right place to do this and a place where it'd be actually very easy for it to work is within the app stores themselves, where my understanding is Apple and

Google, already—or at least Apple, already requires parental consent when a child does a payment with an app. So it should be pretty trivial to pass a law that requires them to make it so that parents have control anytime a child downloads an app and offers consent of that.

And the research that we've done shows that the vast majority of parents want that, and I think that that's the type of legislation, in addition to some of the other ideas that you-all have, that would make this a lot easier for parents.

Senator KLOBUCHAR. Yes. Just to be clear, I remember one mom telling me with all these things she could maybe do that she can't figure out, it's like a faucet overflowing in a sink, and she's out there with a mop while her kids are getting addicted to more and more different apps, and being exposed to material. We've got to make this simpler for parents so they can protect their kids, and I just don't think this is going to be the way to do it.

I think the answer is what Senator Graham has been talking about, which is opening up the halls of the courtroom. So that puts it on you guys to protect these parents, and protect these kids, and then also to pass some of these laws which makes it easier for law enforcement.

Chair DURBIN. Thank you, Senator Klobuchar. We're going to try to stick to the 7-minute rule. Didn't work very well, but we're going to—I'll try to give additional time on the other side as well. Senator, Cornyn.

Senator CORNYN. There's no question that your platforms are very popular, but we know that while here in the United States, we have an open society and a free exchange of information, that there are authoritarian governments, there are criminals who will use your platforms for the sale of drugs, for sex, for extortion, and the like.

And Mr. Chew, I think your company is unique among the ones represented here today because of its ownership by ByteDance, a Chinese company. And I know there have been some steps that you've taken to wall off the data collected here in the United States, but the fact of the matter is that under Chinese law and Chinese National Intelligence Law, all information accumulated by companies in the People's Republic of China are required to be shared with the Chinese Intelligence Services.

ByteDance, the initial release of TikTok, I understand was in 2016. These efforts that you made with Oracle under the so-called Project Texas to wall off the U.S. data was in 2021, and apparently, allegedly, fully walled off in March 2023. What happened to all of the data that TikTok collected before that?

Mr. CHEW. Senator, thank you.

Senator CORNYN. From American users.

Mr. CHEW. I understand. TikTok is owned by ByteDance, which is majority owned by global investors, and we have three Americans on the board out of five. You are right in pointing out that over the last 3 years, we have spent billions of dollars building out Project Texas, which is a plan that is unprecedented in our industry. The wall off, firewall, protected U.S. data from the rest of our staff. We also have this——

Senator CORNYN. And I'm asking about all of the data that you collected prior to that event.

Mr. CHEW. Yes, Senator. We have started a data deletion plan. I talked about this a year ago. We have finished the first phase of data deletion through our data centers outside of the Oracle Cloud Infrastructure. And we're beginning phase two, where we will not only delete from the data centers, we will hire a third party to verify that work. And then we will go into, you know, for example, employees working laptops to delete that as well.

Senator CORNYN. Was all of the data collected by TikTok prior to Project Texas shared with the Chinese government pursuant to the national intelligence laws of that country?

Mr. CHEW. Senator, we have not been asked for any data by the Chinese government, and we have never provided it.

Senator CORNYN. Your company is unique, again, among the ones represented here today because you're currently undergoing review by the Committee on Foreign Investment in the United States. Is that correct?

Mr. CHEW. Senator, yes, there are ongoing discussions, and a lot of our Project Texas work is informed by the discussions with many agencies under the CFIUS umbrella.

Senator CORNYN. Well, CFIUS is designed specifically to review foreign investments in the United States for national security risks. Correct?

Mr. CHEW. Yes, I believe so.

Senator CORNYN. And your company is currently being reviewed by this Interagency Committee at the Treasury Department for potential national security risks?

Mr. CHEW. Senator, this review is on acquisition of Musical.ly, which is an acquisition that was done many years ago.

Senator CORNYN. I mean, is this a casual conversation, or are you actually providing information to the Treasury Department about how your platform operates for evaluating a potential national security risk?

Mr. CHEW. Senator, it's been many years across two administrations, and a lot of discussions around how our plans are, how our systems work. We have a lot of robust discussions about a lot of detail.

Senator CORNYN. Sixty-three percent of teens, I understand, use TikTok. Does that sound about right?

Mr. CHEW. Senator, I cannot verify that. We know we are popular amongst many age groups. The average age in the U.S. today for our user base is over 30, but we are aware we are popular.

Senator CORNYN. And you reside in Singapore with your family. Correct?

Mr. CHEW. Yes. I reside in Singapore, and I work here in the United States as well.

Senator CORNYN. And do your children have access to TikTok in Singapore?

Mr. CHEW. Senator, if they lived in the United States, I would give them access to our under 13 experience. My children are below the age of 13.

Senator CORNYN. My question is, in Singapore, do they have access to TikTok, or is that restricted by domestic law?

Mr. CHEW. We do not have an under 13 experience in Singapore. We have that in the United States because we were deemed a mixed audience app, and we created under 13 experience in response to that.

Senator CORNYN. A Wall Street Journal article published yesterday directly contradicts what your company has stated publicly. According to the journal, employees under the Project Texas say that U.S. user data, including user emails, birth date, IP addresses, continue to be shared with ByteDance staff, again, owned by a Chinese company. Do you dispute that?

Mr. CHEW. Yes, Senator. There are many things about that article. They are inaccurate. Where it gets right is that this is a voluntary project that we built. We spend billions of dollars. There are thousands of employees involved, and it's very difficult because it's unprecedented.

Senator CORNYN. Why is it important that the data collected from U.S. users be stored in the United States?

Mr. CHEW. Senator, this was a project we built in response to some of the concerns that were raised by Members of this Committee and others.

Senator CORNYN. And that was because of concerns that the data that was stored in China could be accessed by the Chinese Communist Party by according to the National Intelligence Law. Correct?

Mr. CHEW. Senator, we are not the only company that does business—you know, that has Chinese employees. For example, we're not even the only company in this room that hires Chinese nationals, but in order to address some of these concerns, we have moved the data into the Oracle Cloud Infrastructure.

We built a 2,000-person team to oversee the management of that data based here. We firewalled it off from the rest of the organization, and then we open it up to third parties like Oracle, and we will onboard others to give them third-party validation. This is unprecedented access. I think we are unique in taking even more steps to protect user data in the United States.

Senator CORNYN. Well, you've disputed The Wall Street Journal story published yesterday. Are you going to conduct any sort of investigation to see whether there's any truth to the allegations made in the article, or are you just going to dismiss them outright?

Mr. CHEW. Oh, we're not going to dismiss them. So we have ongoing security inspections, not only by our own personnel, but also by third parties to ensure that the system is rigorous and robust. No system that any one of us can build is perfect, but what we need to do is to make sure that we are always improving it and testing it against bad people who may try to bypass it. And if anyone breaks our policies within our organization, we will take disciplinary action against them.

Chair DURBIN. Thanks, Senator Cornyn. Senator Coons.

Senator COONS. Thank you, Chairman Durbin. First, I'd like to start by thanking all the families that are here today. All the parents who are here because of a child they have lost. All the families that are here because you want us to see you and to know your concern. You have contacted each of us in our offices expressing your grief, your loss, your passion, and your concern. And the audi-

ence that is watching can't see this, they can see you, the witnesses from the companies, but this room is packed as far as the eye can see.

And when this hearing began, many of you picked up and held pictures of your beloved and lost children. I benefit from and participate in social media, as do many Members of the Committee, and our Nation, and our world. There are now a majority of people on earth participating in and in many ways benefiting from one of the platforms you have launched, or you lead, or you represent.

And we have to recognize there are some real positives to social media. It has transformed modern life, but it has also had huge impacts on families, on children, on nations. And there's a whole series of bills championed by Members of this Committee that tries to deal with the trafficking in illicit drugs, the trafficking in illicit child sexual material, the things that are facilitated on your platforms that may lead to self-harm or suicide.

So we've heard from several of the leaders on this Committee—the Chair, and Ranking, and very talented and experienced Senators. The frame that we are looking at, this is consumer protection. When there is some new technology, we put in place regulations to make sure that it is not overly harmful. As my friend Senator Klobuchar pointed out, one door flew off of one plane, no one was hurt, and yet the entire Boeing fleet of that type of plane was grounded, and a Federal fit-for-purpose agency did an immediate safety review.

I'm going to point not to the other pieces of legislation that I think are urgent that we take up and pass, but to the core question of transparency. If you are a company manufacturing a product that is allegedly addictive and harmful, one of the first things we look to is safety information. We try to give our constituents, our consumers, warnings; labels that help them understand what are the consequences of this product and how to use it safely or not.

As you've heard, pointedly, from some of my colleagues, if you sell an addictive, defective, harmful product in this country in violation of regulations and warnings, you get sued. And what is distinct about platforms as an industry is most of the families who are here, are here because there were not sufficient warnings, and they cannot effectively sue you.

So let me dig in for a moment, if I can, because each of your companies voluntarily discloses information about the content, and the safety investments you make, and the actions you take.

There was a question pressed, I think it was by Senator Graham earlier about TikTok. I believe, Mr. Chew, you said invest $2 billion in safety. My background memo said, your global revenue is $85 billion. Mr. Zuckerberg, my background memo says, you're investing $5 billion in safety in Meta, and your annual revenue is on the order of $116 billion.

So what matters? You can hear some expressions from the parents in the audience. What matters is the relative numbers and the absolute numbers. You are data folks. If there's anybody in this world who understand data, it's you guys. So I want to walk through whether or not these voluntary measures of disclosure of content and harm are sufficient, because I would argue we're here because they're not. Without better information.

How can policymakers know whether the protections you've testified about, the new initiatives, the starting programs, the monitoring, and the takedowns are actually working? How can we understand meaningfully how big these problems are without measuring and reporting data?

Mr. Zuckerberg, your testimony referenced a National Academy of Sciences study that said at the population level, there is no proof about harm for mental health. Well, it may not be at the population level, but I'm looking at a room full of hundreds of parents who have lost children. And our challenge is to take the data and to make good decisions about protecting families and children from harm.

So let me ask about what your companies do or don't report, and I'm going to particularly focus on your content policies around self-harm and suicide. And I'm just going to ask a series of yes or no questions. And what I'm getting at is do you disclose enough.

Mr. Zuckerberg, from your policies prohibiting content about suicide or self-harm, do you report an estimate of the total amount of content, not a percentage of the overall, not a prevalence number, but the total amount of content on your platform that violates this policy? And do you report the total number of views that self-harm or suicide-promoting content that violates this policy gets on your platform?

Mr. ZUCKERBERG. Yes. Senator, we pioneered a quarterly reporting on our community standards enforcement across all these different categories of harmful content. We focus on prevalence, which you mentioned because what we're focused on is what percent of the content that we take down——

Senator COONS. So Mr. Zuckerberg, I'm going to interrupt you.

Mr. ZUCKERBERG [continuing]. Where our systems proactively identify——

Senator COONS. You're very talented. I have very little time left. I'm trying to get an answer to a question, not as a percentage of the total, because remember it's a huge number. So the percentage is small. But do you report the actual amount of content and the amount of views, self-harm content received?

Mr. ZUCKERBERG. No. I believe we focus on prevalence.

Senator COONS. Correct. You don't. Ms. Yaccarino, yes or no. Do you report it or you don't?

Ms. YACCARINO. Senator, as a reminder, we have less than 1 percent of our users that are between the ages of 13 and 17.

Senator COONS. Do you report the absolute number——

Ms. YACCARINO. We report the number of——

Senator COONS [continuing]. Of how many images and how often do you——

Ms. YACCARINO [continuing]. Posts and accounts that we've taken down. In 2023——

Senator COONS. Yes.

Ms. YACCARINO [continuing]. We've taken over almost a million posts down in regards to mental health and self-harm.

Senator COONS. Mr. Chew, do you disclose the number of appearances of these types of content and how many are viewed before they're taken down?

Mr. CHEW. Senator, we disclosed the number we take down based on each category of violation and how many of that were taken down proactively before it was reported.

Senator COONS. Mr. Spiegel.

Mr. SPIEGEL. Yes, Senator, we do disclose.

Senator COONS. Mr. Citron.

Mr. CITRON. Yes, we do.

Senator COONS. So, I've got three more questions I'd love to walk through if I had unlimited time. I will submit them for the record.

The larger point is that platforms need to hand over more content about how the algorithms work, what the content does, and what the consequences are. Not at the aggregate, not at the population level, but the actual numbers of cases so we can understand the content.

In closing, Mr. Chairman, I have a bipartisan bill, the Platform Accountability and Transparency Act, co-sponsored by Senators Cornyn, Klobuchar, Blumenthal on this Committee, and Senator Cassidy and others. It's in front of the Commerce Committee, not this Committee. But it would set reasonable standards for disclosure and transparency to make sure that we're doing our jobs based on data.

Yes, there's a lot of emotion in this field, understandably, but if we're going to legislate responsibly about the management of the content on your platforms, we need to have better data. Is there any one of you willing to say now that you support this bill? Mr. Chairman, let the record reflect a yawning silence from the leaders of the social media platforms. Thank you.

Chair DURBIN. Thanks, Senator Coons. We're on one of two, the first of two roll calls, and so please understand if some of the Members leave and come back. It's no disrespect, they're doing their job. Senator Lee.

Senator LEE. Thank you, Mr. Chairman. Tragically, survivors of sexual abuse are often repeatedly victimized and revictimized over, and over, and over again by having nonconsensual images of themselves on social media platforms. There's a NCMEC study that pointed out there was one instance of CSAM that reappeared more than 490,000 times after it had been reported—after it had been reported.

So we need tools in order to deal with this. We need, frankly, laws in order to mandate standards so that this doesn't happen; so that we have a systematic way of getting rid of this stuff, because there is literally no plausible justification no way of defending this.

One tool, one that I think would be particularly effective is a bill that I'll be introducing later today, and I invite all my Committee Members to join me. It's called the PROTECT Act. The PROTECT Act would, in pertinent part, require websites to verify age and verify that they've received consent of any and all individuals appearing on their site in pornographic images. And it also requires platforms to have meaningful processes for an individual seeking to have images of him or herself removed in a timely manner.

Ms. Yaccarino, based on your understanding of existing law, what might it take for a person to have those images removed, say from X?

Ms. YACCARINO. Senator Lee, thank you. It sounds like what you are going to introduce into law in terms of ecosystem-wide and user consent sounds exactly like part of the philosophy of why we're supporting the SHIELD Act, and no one should have to endure nonconsensual images being shared online.

Senator LEE. Yes. And without that, without laws in place—and it's fantastic anytime a company as you've described with yours, wants to take those steps. It's very helpful. It can take a lot longer than it should, and sometimes it does to the point where somebody had images shared 490,000 times after it was reported to the authorities. And that's deeply concerning. But yes, the PROTECT Act would work in tandem with—it's a good compliment to the SHIELD Act.

Mr. Zuckerberg, let's turn to you next. As you know, I feel strongly about privacy, and believe that one of the best protections for an individual's privacy online involves end-to-end encryption. We also know that a great deal of grooming and sharing of CSAM happens to occur on end-to-end encrypted systems. Tell me, does Meta allow juvenile accounts on its platforms to use encrypted messaging services within those apps?

Mr. ZUCKERBERG. Sorry, Senator, what do you mean juvenile?

Senator LEE. Underage. People under 18.

Mr. ZUCKERBERG. Under 18. We allow people under the age of 18 to use WhatsApp, and we do allow that to be encrypted. Yes.

Senator LEE. Do you have a bottom-level age at which they're not allowed to use it?

Mr. ZUCKERBERG. Yes. I don't think we allow people under the age of 13.

Senator LEE. Okay. What about you, Mr. Citron. Discord, do you allow kids to have accounts to access encrypted messaging?

Mr. CITRON. Discord is not allowed to be used by children under the age of 13, and we do not use end-to-end encryption for text messages. You know, we believe that it's very important to be able to respond to—well, from law enforcement requests, and we're also working on proactively building technology.

We're working with a nonprofit called Thorn to build a grooming classifier so that our Teen Safety Assist feature can actually identify these conversations, if they might be happening, so we can intervene and give those teens tools to get out of that situation, or potentially even report those conversations and those people to law enforcement.

Senator LEE. And then encryption, as much as it can prove useful elsewhere, it can be harmful, especially if you are on a site where, you know, children are being groomed and exploited. If you allow children onto an end-to-end encryption-enabled app that can prove problematic.

Now, let's go back to you for a moment, Mr. Zuckerberg. Instagram recently announced that it's going to restrict all teenagers from access to eating disorder material, suicidal ideation-themed material, self-harm content, and that's fantastic. That's great. What's odd, what I'm trying to understand is why it is that Instagram is only restricting access to sexually explicit content, but only for teens ages 13 to 15. Why not restrict it for 16-and 17-year-olds as well?

Mr. ZUCKERBERG. Senator, my understanding is that we don't allow sexually explicit content on the service for people of any age.

Senator LEE. How is that going?

[Laughter.]

Mr. ZUCKERBERG. You know, our prevalence metrics suggests that, I think, it's 99 percent or so of the content that we remove, we're able to identify automatically using AI systems. So I think that our efforts in this, while they're not perfect, I think are industry-leading.

The other thing that you asked about was self-harm content, which is what we recently restricted, and we made that shift of the—I think the state of the science is shifting a bit. Previously, we believed that when people were thinking about self-harm, it was important for them to be able to express that and get support.

And now more of the thinking in the field is that it's just better to not show that content at all, which is why we recently moved to restrict that from showing up for those teens at all.

Senator LEE. Okay. Is there a way for parents to make a request on what their kid can see or not see on your sites?

Mr. ZUCKERBERG. There are a lot of parental controls. I'm not sure if there—I don't think that we currently have a control around topics, but we do allow parents to control the time that the children are on the site. And also, a lot of it is based on kind of monitoring and understanding what the teen's experience is—what they're interacting with.

Senator LEE. Mr. Citron, Discord allows pornography on its site. Now, reportedly, 17 percent of minors who use Discord have had online sexual interactions on your platform. 17 percent. And 10 percent have those interactions with someone that the minor believed to be an adult. Do you restrict minors from accessing Discord servers that host pornographic material on them?

Mr. CITRON. Senator, yes, we do restrict minors from accessing content that is marked for adults, and Discord also does not recommend content to people. Discord is a chat app, we do not have a feed or an algorithm that boosts content. So, we allow adults to share content with other adults in adult-labeled spaces, and we do not allow teens to access that content.

Senator LEE. Okay. I see my time's expired. Thank you.

Senator WHITEHOUSE [presiding]. Welcome, everyone. We are here in this hearing because, as a collective, your platforms really suck at policing themselves. We hear about it here in Congress with fentanyl and other drug dealing facilitated across platforms. We see it and hear about it here in Congress with harassment and bullying that takes place across your platforms. We see it and hear about it here in Congress with respect to child pornography, sex exploitation, and blackmail, and we are sick of it.

It seems to me that there is a problem with accountability because these conditions continue to persist. In my view, Section 230, which provides immunity from lawsuit, is a very significant part of that problem. If you look at where bullies have been brought to heel recently, whether it's Dominion finally getting justice against Fox News after a long campaign to try to discredit the election equipment manufacturer. Or whether it's the moms and dads of the Sandy Hook victims finally getting justice against InfoWars

and his campaign of trying to get people to believe that the massacre of their children was a fake put on by them; or even now more recently, with a writer getting a very significant judgment against Donald Trump. After years of bullying and defamation, an honest courtroom has proven to be the place where these things get sorted out.

And I'll just describe one case, if I may. It's called *Doe v. Twitter*. The plaintiff in that case was blackmailed in 2017 for sexually explicit photos and videos of himself, then aged 13 to 14. A compilation video of multiple CSAM videos surfaced on Twitter in 2019. A concerned citizen reported that video on December 25, 2019, Christmas Day. Twitter took no action. The plaintiff, then a minor in high school in 2019, became aware of this video from his classmates in January 2020. You're a high school kid, and suddenly there's that. That's a day that's hard to recover from.

Ultimately, he became suicidal. He and his parents contacted law enforcement and Twitter to have these videos removed on January 21, and again on January 22, 2020, and Twitter ultimately took down the video on January 30, 2020, once Federal law enforcement got involved.

That's a pretty foul set of facts. And when the family sued Twitter for all those months of refusing to take down the explicit video of this child, Twitter invoked Section 230, and the district court ruled that the claim was barred.

There is nothing about that set of facts that tells me that Section 230 performed any public service in that regard. I would like to see very substantial adjustments to Section 230 so that the honest courtroom, which brought relief and justice to E. Jean Carroll after months of defamation, which brought silence, peace, and justice to the parents of the Sandy Hook children after months of defamation and bullying by InfoWars and Alex Jones, and which brought significant justice and an end to the campaign of defamation by Fox News to a little company that was busy just making election machines.

So, my time is running out, I'll turn to—I guess Senator Cruz is next, but I would like to have each of your companies put in writing what exemptions from the protection of Section 230 you would be willing to accept, bearing in mind the fact situation in *Doe v. Twitter*, bearing in mind the enormous harm that was done to that young person and that family by the nonresponsiveness of this enormous platform over months, and months, and months, and months.

Again, think of what it's like to be a high school kid, and have that stuff up in the public domain, and have the company that is holding it out there in the public domain react so disinterestedly. Okay? Will you put that down in writing for me? One, two, three, four, five yeses. Done.

Senator WHITEHOUSE. Senator Cruz.

Senator CRUZ. Thank you, Mr. Chairman. Social media is a very powerful tool, but we're here because every parent I know, and I think every parent in America is terrified about the garbage that is directed at our kids. I have two teenagers at home, and the phones they have are portals to predators, to viciousness, to bul-

lying, to self-harm, and each of your companies could do a lot more to prevent it.

Mr. Zuckerberg, in June 2023, The Wall Street Journal reported that Instagram's recommendation systems were actively connecting pedophiles to accounts that were advertising the sale of child sexual abuse material. In many cases, those accounts appear to be run by underage children themselves, often using code words and emojis to advertise illicit material. In other cases, the accounts included indicia that the victim was being sex trafficked.

Now, I know that Instagram has a team that works to prevent the abuse and exploitation of children online, but what was particularly concerning about the Wall Street Journal expose was the degree to which Instagram's own algorithm was promoting the discoverability of victims for pedophiles seeking child abuse material.

In other words, this material wasn't just living on the dark corners of Instagram. Instagram was helping pedophiles find it by promoting graphic hashtags, including #pedowhore and #preteensex, to potential buyers. Instagram also displayed the following warning screen to individuals who were searching for child abuse material, "These results may contain images of child sexual abuse." And then you gave users two choices, "Get resources or see results anyway." Mr. Zuckerberg, what the hell were you thinking?

Mr. ZUCKERBERG. All right. Senator, the basic science behind that is that when people are searching for something that is problematic, it's often helpful to, rather than just blocking it, to help direct them toward something that could be helpful for getting them to get help. But we also——

Senator CRUZ. I understand, "get resources." In what sane universe is there a link for see results anyway?

Mr. ZUCKERBERG. Well, because we might be wrong. We try to trigger this warning, or we try to—when we think that there's any chance that the results might be——

Senator CRUZ. Okay. You might be wrong. Let me ask you, how many times was this warning screen displayed?

Mr. ZUCKERBERG. I don't know, but the——

Senator CRUZ. You don't know. Why don't you know?

Mr. ZUCKERBERG. I don't know the answer to that off the top of my head, but——

Senator CRUZ. You know what, Mr. Zuckerberg, it's interesting you say you don't know it off the top of your head because I asked it in June 2023 in an oversight letter, and your company refused to answer. Will you commit right now to, within 5 days, answering this question for this Committee?

Mr. ZUCKERBERG. We'll follow up on that?

Senator CRUZ. Is that a yes? Not a we'll follow up. I know how lawyers write statements saying we're not going to answer. Will you tell us how many times this warning screen was displayed? Yes or no?

Mr. ZUCKERBERG. Senator, I'll personally look into it. I'm not sure if we have——

Senator CRUZ. Okay. So you're refusing to answer that. Let me ask you this, how many times did an Instagram user who got this warning that you're seeing images of child sexual abuse, how many

times did that user click on, "see results anyway?" I want to see that.

Mr. ZUCKERBERG. Senator, I'm not sure if we stored that, but I'll personally look into this, and we'll follow up after——

Senator CRUZ. And what follow up did Instagram do when you have a potential pedophile clicking on, "I'd like to see child porn." What did you do next when that happened?

Mr. ZUCKERBERG. Senator, I think that an important piece of context here is that any content that we think is child sexual abuse——

Senator CRUZ. Mr. Zuckerberg, that's called a question. What did you do next when someone clicked, "You may be getting child sexual abuse images," and they click, "see results anyway?" What was your next step? You said you might be wrong. Did anyone examine was it in fact child sexual abuse material? Did anyone report that user? Did anyone go and try to protect that child? What did you do next?

Mr. ZUCKERBERG. Senator, we take down anything that we think is sexual abuse material on the service, and we do——

Senator CRUZ. Did anyone verify whether it was in fact child sexual abuse material?

Mr. ZUCKERBERG. Senator, I don't know if every single search result we're following up on, but in——

Senator CRUZ. Did you report the people who wanted it?

Mr. ZUCKERBERG. Senator, do you want me to answer your question?

Senator CRUZ. Yes. I want you to answer the question I'm asking. Did you report——

Mr. ZUCKERBERG. Give me some time to speak then.

Senator CRUZ [continuing]. The people who click, "see results anyway?"

Mr. ZUCKERBERG. That's probably one of the factors that we use in reporting, and in general, and we've reported more people and done more reports like this to NCMEC, the National Center of Missing Exploited Children, than any other company in the industry. We proactively go out of our way across our services to do this, and have made it—I think, it's more than 26 million reports, which is more than the whole rest of the industry combined. So I think the allegation——

Senator CRUZ. So Mr. Zuckerberg——

Mr. ZUCKERBERG [continuing]. That we don't take this seriously——

Senator CRUZ [continuing]. Your company and every social media company needs to do much more to protect children. All right. Mr. Chew, in the next couple of minutes I have, I want to turn to you. Are you familiar with China's 2017 National Intelligence Law, which states, "All organizations and citizens shall support, assist, and cooperate with national intelligence efforts in accordance with the law, and shall protect national intelligence work secrets they are aware of?"

Mr. CHEW. Yes. I'm familiar with this.

Senator CRUZ. TikTok is owned by ByteDance. Is ByteDance subject to the law?

Mr. CHEW. For the Chinese businesses that ByteDance owns, yes, it will be subject to this, but TikTok is not available in Mainland China. And Senator, as we talked about in your office, we built Project Texas to put this out of reach.

Senator CRUZ. So, ByteDance is subject to the law. Now, under this law, which says, "shall protect national intelligence work secrets they're aware of," it compels people subject to the law to lie to protect those secrets. Is that correct?

Mr. CHEW. I cannot comment on that. What I said, again, is that we have——

Senator CRUZ. Because you have to protect those secrets.

Mr. CHEW. No, Senator, TikTok is not available in Mainland China. We have moved the data into an American product infrastructure——

Senator CRUZ. But TikTok is controlled by ByteDance, which is subject to this law. Now, you said earlier, and I wrote this down, "We have not been asked for any data by the Chinese government, and we have never provided it." I'm going to tell you, and I told you this when you and I met last week in my office, I do not believe you, and I'll tell you, the American people don't either.

If you look at what is on TikTok in China, you are promoting to kids' science and math videos, educational videos, and you limit the amount of time kids can be on TikTok. In the United States, you are promoting to kids' self-harm videos and anti-Israel propaganda. Why is there such a dramatic difference?

Mr. CHEW. Senator, that is just not accurate. There is a lot of——

Senator CRUZ. There's not a difference between what kids see in China and what kids see here?

Mr. CHEW. Senator, TikTok is not available in China. It's a separate experience there. But what I'm saying is——

Senator CRUZ. But you have a company that is essentially the same except it promotes beneficial materials instead of harmful materials.

Mr. CHEW. That is not true. We have a lot of science and math content here on TikTok. There's so much of it——

Senator CRUZ. All right. Let me point to this, Mr. Chew. There was a report recently that compared hashtags on Instagram to hashtags on TikTok, and what trended, and the differences were striking. So for something like #TaylorSwift or #Trump, researchers found roughly two Instagram posts for every one on TikTok. That's not a dramatic difference.

That difference jumps to 8-to-1 for the #Uyghur, and it jumps to 30-to-1 for the #Tibet, and it jumps to 57-to-1 for #Tiananmen Square, and it jumps to 174-to-1 for the #HongKongProtest. Why is it that on Instagram people can put up a #HongKongProtest 174 times compared to TikTok? What censorship is TikTok doing at the request of the Chinese government?

Mr. CHEW. None. Senator——

Senator CRUZ. Can you explain that differential?

Mr. CHEW. That analysis is flawed, has been debunked by other external sources like the Cato Institute. Fundamentally, a few things happen here. Not all videos carry hashtags. That's the first

thing. The second thing is that you cannot selectively choose a few words within a certain time period——

Senator CRUZ. Why the difference between Taylor Swift and Tiananmen Square? What happened at Tiananmen Square?

Mr. CHEW. Senator, there was a massive protest during that time. But what I'm trying to say is our users can freely come and post this content——

Senator CRUZ. Why would there be no difference on Taylor Swift or a minimal difference, and a massive difference on Tiananmen Square, Hong Kong?

Chair DURBIN [presiding]. Senator, could you wrap up, please?

Mr. CHEW. Senator, our algorithm does not suppress a new content simply based on——

Senator CRUZ. Could you answer that question? Why is there a difference?

Mr. CHEW. Like I said, I think this analysis is flawed. You're selectively choosing some words over some periods. We haven't been around this——

Senator CRUZ. There is an obvious——

Mr. CHEW [continuing]. And other apps——

Senator CRUZ [continuing]. Difference. 174-to-1 for Hong Kong compared to Taylor Swift is dramatic.

Chair DURBIN. Senator Blumenthal.

Senator BLUMENTHAL. Mr. Zuckerberg, you know who Antigone Davis is, correct?

Mr. ZUCKERBERG. Yes.

Senator BLUMENTHAL. She's one of your top leaders. In September 2021, she was global head of safety, correct?

Mr. ZUCKERBERG. Yes.

Senator BLUMENTHAL. And you know that she came before a Subcommittee, the Commerce Committee that I chaired at the time, Subcommittee on Consumer Protection, correct?

Mr. ZUCKERBERG. Yes.

Senator BLUMENTHAL. And she was testifying on behalf of Facebook, right?

Mr. ZUCKERBERG. Meta, but, yes.

Senator BLUMENTHAL. It was then Facebook, but Meta now. And she told us, and I'm quoting, "Facebook is committed to building better products for young people and to doing everything we can to protect their privacy, safety, and well-being on our platforms."

And she also said kids' safety is an area where, "we are investing heavily." We now know that statement was untrue. We know it from an internal email that we have received. It's an email written by Nick Clegg. You know who he is, correct?

Mr. ZUCKERBERG. Yes.

Senator BLUMENTHAL. He was Meta's president of global affairs, and he wrote a memo to you which you received, correct? It was written to you.

Mr. ZUCKERBERG. Senator, I can't see the email, but sure, I'll assume that you got it. Correct.

Senator BLUMENTHAL. And he summarized Facebook's problems. He said, "We are not on track to succeed for our core well-being topics; problematic use, bullying and harassment connections, and SSI," meaning suicidal self-injury. He said also in another memo,

"We need to do more, and we are being held back by a lack of investment." This memo has the date of August 28, just weeks before that testimony from Antigone Davis. Correct?

Mr. ZUCKERBERG. Sorry, Senator, I'm not sure what the date the testimony was.

Senator BLUMENTHAL. Well, those are the dates on the emails. Nick Clegg was asking you, pleading with you for resources to back up the narrative to fulfill the commitments. In effect, Antigone Davis was making promises that Nick Clegg was trying to fulfill, and you rejected that request for 45 to 84 engineers to do well-being or safety.

We know that you rejected it from another memo. Nick Clegg's assistant, Tim Colburne, who said Nick did email Mark referring to that earlier email to emphasize his support for the package, but it lost out to the various other pressures and priorities.

We've done a calculation that those, potentially, 84 engineers would've cost Meta about $50 million in a quarter when it earned $9.2 billion. And yet it failed to make that commitment in real terms, and you rejected that request because of other pressures and priorities. That is an example from your own internal document of failing to act. And it is the reason why we can no longer trust Meta, and, frankly, any of the other social media to in effect grade their own homework.

The public, and particularly the parents in this room, know that we can no longer rely on social media to provide the kind of safeguards that children and parents deserve. And that is the reason why passing the Kids Online Safety Act is so critically important.

Mr. Zuckerberg, do you believe that you have a constitutional right to lie to Congress?

Mr. ZUCKERBERG. Senator, no, but I mean you——

Senator BLUMENTHAL. Well, let me just clarify for you.

Mr. ZUCKERBERG [continuing]. Quoted a bunch of words, and I'd like the opportunity to respond to——

Senator BLUMENTHAL. Let me just clarify for you. In a lawsuit brought by hundreds of parents, some in this very room, alleging that you made false and misleading statements concerning the safety of your platform for children. You argued in not just one pleading, but twice, in December, and then in January, that you have a constitutional right to lie to Congress. Do you disavow that filing in court?

Mr. ZUCKERBERG. Senator, I don't know what filing you're talking about, but I testified——

Senator BLUMENTHAL. It's a filing from——

Mr. ZUCKERBERG [continuing]. Honestly and truthfully, and I would like the opportunity to respond to the previous things that you showed as well.

Senator BLUMENTHAL. Well, I have a few more questions, and let me ask others who are here because I think it's important to put you on record. Who will support the Kids Online Safety Act? Yes or no. Mr. Citron?

Mr. CITRON. There are parts of the Act that we think are great and——

Senator BLUMENTHAL. No. It's a yes or no question. I'm going to be running out of time. So I'm assuming the answer is no if you can't answer yes.

Mr. CITRON. We very much think that the National——

Senator BLUMENTHAL. That's a no.

Mr. CITRON [continuing]. Privacy Standard would be great.

Senator BLUMENTHAL. Mr. Siegel.

Mr. SIEGEL. Senator, we strongly support the Kids' Online Safety Act, and we've already implemented many of its core provisions.

Senator BLUMENTHAL. Thank you. I welcome that support along with Microsoft's support. Mr. Chew.

Mr. CHEW. Senator, with some changes we can support it.

Senator BLUMENTHAL. Now in its present form, do you support it? Yes, or no?

Mr. CHEW. We are aware that some groups have raised some concerns. It's important to understand how——

Senator BLUMENTHAL. I'll take that as a no. Ms. Yaccarino.

Ms. YACCARINO. Senator, we support KOSA, and will continue to make sure that it accelerates, and make sure it continues to offer a community for teens that are seeking that voice.

Senator BLUMENTHAL. Mr. Zuckerberg.

Mr. ZUCKERBERG. Senator, we support the age-appropriate content standards, but would have some suggestions——

Senator BLUMENTHAL. Yes or no, Mr. Zuckerberg.

Mr. ZUCKERBERG [continuing]. On how to implement it.

Senator BLUMENTHAL. Do you support the Kids Online Safety Act?

Mr. ZUCKERBERG. Senator, I think these are nuanced——

Senator BLUMENTHAL. You're in public, and I'm just asking whether you'll support it or not.

Mr. ZUCKERBERG. These are nuanced things. I think that the basic spirit is right. I think the basic ideas in it are right, and there are some ideas that I would debate how to best——

Senator BLUMENTHAL. Unfortunately, I don't think we can count on social media, as a group, or Big Tech, to support this measure. And in the past, we know it's been opposed by armies of lawyers and lobbyists. We're prepared for this fight.

But I am very, very glad that we have parents here because tomorrow we're going to have an advocacy day, and the folks who really count, the people in this room who support this measure, are going to be going to their representatives and their Senators, and their voices and faces are going to make a difference.

Senator Schumer has committed that he will work with me to bring this bill to a vote, and then we will have real protection for children and parents online. Thank you, Mr. Chairman.

Chair DURBIN. Thank you, Senator Blumenthal. We have a vote on. Has Senator Cotton—have you voted and Senator Hawley. You haven't voted yet? You're next. And I don't know how long the vote will be open, but I'll turn it over to you.

Senator HAWLEY. Thank you, Mr. Chairman. Mr. Zuckerberg, let me start with you. Did I hear you say in your opening statement that there's no link between mental health and social media use?

Mr. ZUCKERBERG. Senator, what I said is, I think it's important to look at the science. I know people widely talk about this as if

that is something that's already been proven, and I think that the bulk of the scientific evidence does not support that.

Senator HAWLEY. Well, really, let me just remind you of some of the science from your own company. Instagram studied the effect of your platform on teenagers. Let me just read you some quotes from The Wall Street Journal's report on this, company "Researchers found that Instagram is harmful for a sizable percentage of teenagers, most notably teenage girls."

Here's a quote from your own study. "We make body image issues worse for 1-in-3 teen girls." Here's another quote. "Teens blamed Instagram—" this is your study, "for increases in the rate of anxiety and depression. This reaction was unprompted and consistent across all groups." That's your study.

Mr. ZUCKERBERG. Senator, we try to understand the feedback and how people feel about the services. We can improve——

Senator HAWLEY. Wait a minute, your own study says that you make life worse for one in three teenage girls. You increase——

Mr. ZUCKERBERG. No, Senator, that's not what it says.

Senator HAWLEY [continuing]. Anxiety and depression. That's what it says, and you're here testifying to us in public that there's no link. You've been doing this for years. For years you've been coming in public and testifying under oath that there's absolutely no link, your product is wonderful, the science is nascent, full speed ahead, while internally, you know full well your product is a disaster for teenagers.

Mr. ZUCKERBERG. Senator, that's not true.

Senator HAWLEY. And you keep right on doing what you're doing, right?

[Applause.]

Mr. ZUCKERBERG. That's not true. That's not true.

Senator HAWLEY. Let me show you some other facts——

Mr. ZUCKERBERG. We can show you data if you want——

Senator HAWLEY [continuing]. I know that you're familiar with—wait a minute, wait a minute. That's not a question. That's not a question. Those are facts, Mr. Zuckerberg. That's not a question.

Mr. ZUCKERBERG. Those aren't facts.

Senator HAWLEY. Let me show you some more facts. Here's some information from a whistleblower who came before the Senate, testified under oath in public. He worked for you. It's a senior executive. Here's what he showed he found when he studied your products.

So for example, this is girls between the ages of 13 and 15 years old. Thirty-seven percent of them reported that they had been exposed to nudity on the platform, unwanted, in the last 7 days. Twenty-four percent said that they had experienced unwanted sexual advances. They'd been propositioned in the last 7 days. Seventeen percent said they had encountered self-harm content pushed at them in the last 7 days.

Now, I know you're familiar with these stats because he sent you an email where he lined it all out. I mean, we've got a copy of it right here. My question is, who did you fire for this? Who got fired because of that?

Mr. ZUCKERBERG. Senator, we study all this because it's important and we want to improve our services.

Senator HAWLEY. Well, you just told me a second ago you studied it. That there was no linkage. Who did you fire?

Mr. ZUCKERBERG. Senator, I said you mischaracterized——

Senator HAWLEY. Thirty-seven percent of teenage girls between 13 and 15 were exposed to unwanted nudity in a week on Instagram. You knew about it. Who did you fire?

Mr. ZUCKERBERG. Senator, this is why we're building all——

Senator HAWLEY. Who did you fire?

Mr. ZUCKERBERG. Senator, I don't think that that's——

Senator HAWLEY. Who did you fire?

Mr. ZUCKERBERG. I'm not going to answer that.

Senator HAWLEY. It's because you didn't fire anybody, right? You didn't——

Mr. ZUCKERBERG. Senator, I don't think——

Senator HAWLEY [continuing]. Take any significant action.

Mr. ZUCKERBERG [continuing]. It's not appropriate to talk about, like, H.R. decisions——

Senator HAWLEY. It's not appropriate? Do you know who's sitting behind you? You've got families from across the Nation whose children are either severely harmed, or gone, and you don't think it's appropriate to talk about steps that you took, the fact that you didn't fire a single person? Let me ask you this. Have you compensated any of the victims?

Mr. ZUCKERBERG. Sorry?

Senator HAWLEY. Have you compensated any of the victims? These girls, have you compensated them?

Mr. ZUCKERBERG. I don't believe so.

Senator HAWLEY. Why not? Don't you think they deserve some compensation for what your platform has done? Help with counseling services, help with dealing with the issues that your services caused?

Mr. ZUCKERBERG. Our job is to make sure that we build tools to help keep people safe.

Senator HAWLEY. Are you going to compensate them?

Mr. ZUCKERBERG. Senator, our job, and what we take seriously is making sure that we build industry-leading tools to find harmful content——

Senator HAWLEY. To make money.

Mr. ZUCKERBERG [continuing]. And take it off the services——

Senator HAWLEY. To make money.

Mr. ZUCKERBERG [continuing]. And to build tools that empower parents.

Senator HAWLEY. So you didn't take any action. You——

Mr. ZUCKERBERG. That's not true, Senator.

Senator HAWLEY [continuing]. Didn't fire anybody. You haven't compensated a single victim. Let me ask you this. There's families of victims here today. Have you apologized to the victims?

Mr. ZUCKERBERG. I——

Senator HAWLEY. Would you like to do so now?

Mr. ZUCKERBERG. Well——

Senator HAWLEY. They're here. You're on national television. Would you like now to apologize to the victims who have been harmed by your product? Show him the pictures.

[Applause.]

Senator HAWLEY. Would you like to apologize for what you've done to these good people?

[Addressing audience.]

Mr. ZUCKERBERG. I'm sorry for everything that you've all gone through. It's terrible. No one should have to go through the things that your families have suffered. And this is why we invest so much, and are going to continue doing industry-leading efforts to make sure that no one has to go through the types of things that your families have had to suffer.

Senator HAWLEY. You know, why Mr. Zuckerberg, why should your company not be sued for this? Why is it that you can claim—you hide behind a liability shield? You can't be held accountable. Shouldn't you be held accountable personally? Will you take personal responsibility?

Mr. ZUCKERBERG. Senator, I think I've already answered this. I mean, these issues——

Senator HAWLEY. We'll try this again. Will you take personal responsibility?

Mr. ZUCKERBERG. Senator, I view my job and the job of our company as building the best tools that we can to keep our community safe——

Senator HAWLEY. Well, you're failing at that.

Mr. ZUCKERBERG. Well, Senator, we're doing an industry-leading effort. We build AI tools that——

Senator HAWLEY. Oh, nonsense. Your product is killing people. Will you personally commit to compensating the victims? You're a billionaire. Will you commit to compensating the victims? Will you set up a compensation fund——

Mr. ZUCKERBERG. Senator——

Senator HAWLEY [continuing]. With your money?

Mr. ZUCKERBERG [continuing]. I think these are complicated——

Senator HAWLEY. With your money.

Mr. ZUCKERBERG. Senator, these are complicated issues——

Senator HAWLEY. No, that's not a complicated question. That's a yes, or no. Will you set up a victim's compensation fund with your money, the money you made on these families sitting behind you? Yes, or no?

Mr. ZUCKERBERG. Senator, I don't think that that's—my job——

Senator HAWLEY. Sounds like a no.

Mr. ZUCKERBERG [continuing]. Is to make sure we make good tools. My job is to make sure——

Senator HAWLEY. Sounds like a no. Your job is to be responsible for what your company has done. You've made billions of dollars on the people sitting behind you here. You've done nothing to help them. You've done nothing to compensate them. You've done nothing to put it right. You could do so here today and you should. You should, Mr. Zuckerberg.

Before my time expires, Mr. Chew, let me just ask you. Your platform, why should your platform not be banned in the United States of America? You are owned by a Chinese communist company or a company based in China. The editor-in-chief of your parent company is a Communist Party Secretary. Your company has been surveilling Americans for years.

According to leaked audio from more than 80 internal TikTok meetings, China-based employees of your company have repeatedly accessed nonpublic data of United States citizens. Your company has tracked journalists, improperly gaining access to their IP addresses user data in an attempt to identify whether they're writing negative stories about you. Why should—your platform is basically an espionage arm for the Chinese Communist Party. Why should you not be banned in the United States of America?

Mr. CHEW. Senator, I disagree with your characterization. Many of what you have said, we have explained in a lot of detail. TikTok is used by 170 million Americans.

Senator HAWLEY. I know, but when every single one of those Americans are in danger from the fact that you track their keystrokes, you track their app usage, you track their location data, and we know that all of that information can be accessed by Chinese employees who are subject to the dictates of the Chinese Communist Party.

Mr. CHEW. That is not——

Senator HAWLEY. Why should you not be banned in this country?

Mr. CHEW. Senator, that is not accurate. A lot of what you described we collect, we don't.

Senator HAWLEY. It is 100 percent accurate. Do you deny that, repeatedly, American's data has been accessed by ByteDance employees in China?

Mr. CHEW. We built a project that cost us billions of dollars to stop that, and we have made a lot of progress——

Senator HAWLEY. And it hasn't been stopped. According to The Wall Street Journal report from just yesterday, even now, "ByteDance workers, without going through official channels, have access to the private information of American citizen"—I'm quoting from the article—"private information of American citizens, including their birthday, their IP address, and more." That's now.

Mr. CHEW. Senator, as we know, the media doesn't always get it right. What we have, what we have——

Senator HAWLEY. But the Chinese Communist Party does?

Mr. CHEW. I'm not saying that. What I'm saying is that we have been—we have spent billions of dollars to build this project. It's rigorous, it's robust, it's unprecedented, and I'm proud of the work that the 2,000 employees are doing to protect the data of American users.

Senator HAWLEY. But it's not protected. That's the problem, Mr. Chew. It's not protected at all. It's subject to Communist Chinese Party inspection and review, your app, unlike anybody else sitting here, and heaven knows I've got problems with everybody here. But your app, unlike any of those, is subject to the control and inspection of a foreign hostile government that is actively trying to track the information of whereabouts of every American that they get their hands on. Your app ought to be banned in the United States of America for the security of this country.

[Applause.]

Senator HAWLEY. Thank you, Mr. Chairman.

Chair DURBIN. Senator Hirono.

Senator HIRONO. Thank you, Mr. Chairman. As we've heard, children face all sorts of dangers when they use social media from

mental health harms to sexual exploitation, even trafficking. Sex trafficking is a serious problem in my home State of Hawaii, especially for native Hawaiian victims. Social media platforms are being used to facilitate this trafficking, as a well as the creation and distribution of CSAM is deeply concerning. But it's happening.

For example, several years ago, a military police officer stationed in Hawaii was sentenced to 15 years in prison for producing CSAM. As part of his online exploitation of a minor female, he began communicating with this 12-year-old girl through Instagram. He then used Snapchat to send her sexually explicit photos and to solicit such photos from her. He later used these photos to blackmail her.

And just last month the FBI arrested a neo-Nazi cult leader in Hawaii who lured victims to his Discord server. He used that server to share images of extremely disturbing child sexual abuse material interspersed with Neo-Nazi imagery. Members of his child exploitation and hate group are also present on Instagram, Snapchat, X, and TikTok, all of which they used to recruit potential members and victims.

In many cases, including the ones I just mentioned, your companies played a role in helping law enforcement investigate these offenders. But by the time of the investigation, so much damage had already been done.

This hearing is about how to keep children safe online, and we've listened to all of your testimony to seemingly impressive safeguards for young users. You try to limit the time that they spend, you require parental consent, you have all of these tools. Yet, trafficking and exploitation of minors online and on your platforms continues to be rampant.

Nearly all of your companies make your money through advertising, specifically by selling the attention of your users. Your product is your users. As a made-up product designer wrote in an email, "Young ones are the best ones. You want to bring people to your service young and early." In other words, hook them early.

Research published last month by Harvard School of Public Health estimates that Snap makes an astounding 41 percent of its revenues by addressing to users under 18. With TikTok, it's 35 percent. Seven of the 10 largest Discord servers attracting many paying users are for games used primarily by teens, by children.

All this is to say that social media companies, yours, and others, make money by attracting kids to your platforms. But ensuring safety doesn't make money. It costs money. If you are going to continue to attract kids to your platforms, you have an obligation to ensure they're safe on the platforms because the current situation is untenable. That is why we're having this hearing. But to ensure safety for our children, that costs money. Your companies cannot continue to profit off young users only to look the other way when those users, our children, are harmed online.

We've had a lot of comments about Section 230 protections, and I think we are definitely heading in that direction. And some of the five bills that we have already passed out of this Committee talks about limiting the liability protections for you.

Senator HIRONO. This is for Mr. Zuckerberg. Last November, the Privacy and Technology Subcommittee heard testimony from

Arturo Béjar. In response to one of my questions about how to ensure that social media companies focus more on child safety, he said, and I am paraphrasing a little bit, Mr. Béjar said what will change their behavior is at the moment that Mark Zuckerberg declares earnings, and these earnings have to be declared to the SEC.

So he has to say, last quarter we made $34 billion, and the next thing he has to say is how many teens experienced unwanted sexual advances on his platform. Mr. Zuckerberg, will you commit to reporting measurable child safety data on your quarterly earnings reports and calls?

Mr. ZUCKERBERG. Senator, it's a good question. We actually already have a quarterly report that we issue and do a call to answer questions for how we're enforcing our community standards. That includes not just the child safety issues and metrics——

Senator HIRONO. So is that a yes?

Mr. ZUCKERBERG. We have a separate call that we do this on, but we've led the industry——

Senator HIRONO. I think that, you know, you have to report your earnings to the SEC. Will you report to them this kind of data—and by numbers, by the way, because as Senator Coons said and others have said percentages don't really tell the full story. Will you report to the SEC the number of teens—and sometimes you don't even know whether they're teens or not, because they just claim to be adults.

Will you report the number of underage children on your platforms who experience unwanted CSAM and other kinds of messaging that harm them? Will you commit to citing those numbers to the SEC when you make your quarterly report?

Mr. ZUCKERBERG. Well, Senator, I'm not sure it would make as much sense to include in the SEC filing, but we file it publicly so that way everyone can see this. And I'd be happy to follow up and talk about what specific metrics. I think the specific things are some of the ones that you just mentioned around underage people under our services, we don't allow people under the age of 13 on our service. So if we find anyone who's under the age of 13, we remove them from our service. Now, I'm not saying that people don't lie and that there aren't——

Senator HIRONO. Yes, apparently, they're.

Mr. ZUCKERBERG [continuing]. Anyone who's under the age of 13 who's using it, but I'm not going to be able to—we're not going to be able to count how many people there are because, fundamentally, if we identify that someone is underage, we've removed them from the service.

Senator HIRONO. I think that's really important that we get actual numbers because these are real human beings. That's why all these parents and others are here. Because each time that a young person is exposed to this kind of unwanted material and they get hooked, it is a danger to that individual. So, I'm hoping that you are saying that you do report this kind of information to, if not the SEC, that it is made public. I think I'm hearing that yes you do, so.

Mr. ZUCKERBERG. Yes, Senator. I think we report more publicly on our enforcement than any other company in the industry, and we're very supportive of transparency measures.

Senator HIRONO. I'm running out of time, Mr. Zuckerberg, but so I will follow up with what exactly it is that you do report.

Senator HIRONO. Again, for you, when Meta automatically places young people's accounts—and you testified to this—on the most restrictive privacy and content sensitivity sessions, and yet teens are able to opt out of these safeguards. Isn't that right?

Mr. ZUCKERBERG. Yes.

Senator HIRONO. It's not mandatory that they remain on these settings. They can opt out.

Mr. ZUCKERBERG. Senator, yes, we default teens into a private account. So they have a private and restricted experience, but some teens want to be creators, and want to have content that they share more broadly. And I don't think that that's something that should just blanketly be banned.

Senator HIRONO. Why not? I think it should be mandatory that they're not—that they remain on the more restrictive settings.

Mr. ZUCKERBERG. Senator, I think there's——

Senator HIRONO. They have to start somewhere.

Mr. ZUCKERBERG. I mean, a lot of teens create amazing things, and I think with the right supervision, and parenting, and controls, I think that that's like—I don't think that that's the type of thing that you want to just not allow anyone to be able to do. I think you want to make it so that——

Senator HIRONO. My time is up, but I have to say that there is an argument that you-all make for every single thing that we are proposing. And I share the concern that I have about the blanket limitation on liabilities that we provide all of you. And I think that that has to change, and that is on us, on Congress, to make that change. Thank you, Mr. Chairman.

Chair DURBIN. Thank you, Senator Hirono. Senator Cotton.

Senator COTTON. Mr. Chew, let's cut straight to the chase. Is TikTok under the influence of the Chinese Communist Party?

Mr. CHEW. No, Senator. We are a private business.

Senator COTTON. Okay. So you can say that your parent, ByteDance, is subject to the 2017 National Security Law, which requires Chinese companies to turn over information to the Chinese government and conceal it from the rest of the world. You concede that, correct?

Mr. CHEW. Senator, the Chinese business——

Senator COTTON. There's no question, you conceded it earlier.

Mr. CHEW. Any global businesses that does business in China has to follow the local laws.

Senator COTTON. Okay. Isn't it the case that ByteDance also has an internal Chinese Communist Party committee?

Mr. CHEW. Like I said, all businesses that operate in China have developed their local law.

Senator COTTON. So your parent company is subject to the National Security law that requires it to answer the party. It has its own internal Chinese Communist Party committee. You answer to that parent company, but you expect us to believe that you're not under the influence of the Chinese Communist Party?

Mr. CHEW. I understand this concern, Senator, which is why we built Project Texas.

Senator COTTON. It was a yes or no question. Okay. But you used to work for ByteDance, didn't you? You were the CFO for ByteDance?

Mr. CHEW. That is correct, Senator.

Senator COTTON. In April, 2021, while you were the CFO, the Chinese Communist Party's China Internet Investment Fund purchased a 1 percent stake in ByteDance's main Chinese subsidiary, the ByteDance Technology Company. In return for that so-called 1 percent golden share, the party took one of three board seats at that subsidiary company. That's correct, isn't it?

Mr. CHEW. It's for the Chinese business.

Senator COTTON. Is that correct?

Mr. CHEW. It is for the Chinese business.

Senator COTTON. Yes. That deal was finalized on April 30, 2021. Isn't it true that you were appointed the CEO of TikTok on the very next day, on May 1, 2021?

Mr. CHEW. Well, it is a coincidence.

Senator COTTON. It's a coincidence——

Mr. CHEW. Yes.

Senator COTTON [continuing]. That you were the CFO——

Mr. CHEW. Senator, that——

Senator COTTON [continuing]. And then the Chinese Communist Party took its golden share in its board seat, and the very next day you were appointed the CEO of TikTok. That's a hell of a coincidence.

Mr. CHEW. It really is, Senator.

Senator COTTON. Yes, it is. Okay. And before ByteDance, you were at a Chinese company called Xiaomi. Is that correct?

Mr. CHEW. Yes. I used to work around the world.

Senator COTTON. Where did you live when you worked at Xiaomi?

Mr. CHEW. I lived in China. There were many experts.

Senator COTTON. Where exactly?

Mr. CHEW. In Beijing, in China.

Senator COTTON. How many years did you live in Beijing?

Mr. CHEW. Senator, I worked there for about 5 years.

Senator COTTON. So you lived there for 5 years?

Mr. CHEW. Yes.

Senator COTTON. Is it the case that Xiaomi was sanctioned by the U.S. Government in 2021 for being a Communist Chinese military company?

Mr. CHEW. I'm here to talk about TikTok. I think—I think they then had a lawsuit and it was overturned. I can't remember the details.

Senator COTTON. No, no——

Mr. CHEW. It's another company.

Senator COTTON [continuing]. It's the Biden administration that reversed those sanctions just like—by the way, they reversed the terrorist designation on the Houthis in Yemen. How's that working out for them? But it was sanctioned as a Chinese Communist military company. So you said today, as you often say, that you live in Singapore. Of what nation are you a citizen?

Mr. CHEW. Singapore.

Senator COTTON. Are you a citizen of any other nation?

Mr. CHEW. No, Senator.

Senator COTTON. Have you ever applied for Chinese citizenship?

Mr. CHEW. Senator, I served my Nation in Singapore. No, I did not.

Senator COTTON. Do you have a Singaporean passport?

Mr. CHEW. Yes. And I served my military for 2½ years in Singapore.

Senator COTTON. Do you have any other passports from——

Mr. CHEW. No, Senator.

Senator COTTON [continuing]. Any other nations? Your wife is an American citizen, your children are American citizens?

Mr. CHEW. That's correct.

Senator COTTON. Have you ever applied for American citizenship?

Mr. CHEW. No, not yet.

Senator COTTON. Okay. Have you ever been a member of the Chinese Communist Party?

Mr. CHEW. Senator, I'm Singaporean. No.

Senator COTTON. Have you ever been associated or affiliated with the Chinese Communist Party?

Mr. CHEW. No, Senator. Again, I'm Singaporean.

Senator COTTON. Let me ask you something, hopefully a simple question. You said earlier in response to a question that what happened at Tiananmen Square in June 1989 was a massive protest. Anything else happen in Tiananmen Square?

Mr. CHEW. Yes, I think it's well documented. It was a massacre there. Yes.

Senator COTTON. There was an indiscriminate slaughter of hundreds or thousands of Chinese citizens. Do you agree with the Trump administration and the Biden administration, that the Chinese government is committing genocide against the Uyghur people?

Mr. CHEW. Senator, I've said this before. I think it's really important that anyone who cares about this topic or any topic can freely express themselves on TikTok.

Senator COTTON. It's a very simple question that unites both parties in our country and governments around the world. Is the Chinese government committing genocide against the Uyghur people?

Mr. CHEW. Senator, anyone, including, you know, you, can come onto TikTok——

Senator COTTON. Yes or no——

Mr. CHEW [continuing]. And talk about this topic——

Senator COTTON [continuing]. I'm asking you.

Mr. CHEW [continuing]. Or any topic you don't understand.

Senator COTTON. You're a worldly, cosmopolitan, well-educated man who's expressed many opinions on many topics. Is the Chinese government committing genocide against the Uyghur people?

Mr. CHEW. Actually, Senator, I talk mainly about my company——

Senator COTTON. Yes, or no?

Mr. CHEW [continuing]. And I'm here to talk about what TikTok does.

Senator COTTON. Yes, or no?

Mr. CHEW. We allow——

Senator COTTON. You're here to give testimony—to give testimony that's truthful, and honest, and complete. Let me ask you this. Joe Biden last year said that Xi Jinping was a dictator. Do you agree with Joe Biden? Is Xi Jinping a dictator?

Mr. CHEW. Senator, I'm not going to comment on any world leaders.

Senator COTTON. Why won't you answer these very simple questions?

Mr. CHEW. Senator, it's not appropriate for me as a businessman to comment on the world leaders.

Senator COTTON. Are you scared that you'll lose your job if you say anything negative about the Chinese Communist Party?

Mr. CHEW. I disagree with that. You'll find content that is critical of China on our platform.

Senator COTTON. The next time you go on—are you scared that you'll be arrested and disappear the next time you go to Mainland China?

Mr. CHEW. Senator, you will find content that's critical of China and any other country freely on TikTok.

Senator COTTON. Okay. Let's turn to what TikTok, a tool of the Chinese Communist Party, is doing to America's youth. Does the name Mason Edens ring a bell?

Mr. CHEW. Senator, you may have to give me more specifics, if you don't mind.

Senator COTTON. Yes. He was a 16-year-old Arkansan. After a breakup in 2022, he went on your platform and searched for things like inspirational quotes and positive affirmations. Instead, he was served up numerous videos glamorizing suicide until he killed himself by gun. What about the name Chase Nasca? Does that ring a bell?

Mr. CHEW. Would you mind giving me more details, please?

Senator COTTON. He was a 16-year-old who saw more than 1,000 videos on your platform about violence and suicide until he took his own life by stepping in front of a train. Are you aware that his parents, Dean and Michelle, are suing TikTok and ByteDance for pushing their son to take his own life?

Mr. CHEW. Yes, I'm aware of that.

Senator COTTON. Okay. Finally, Mr. Chew, has the Federal Trade Commission sued TikTok during the Biden administration?

Mr. CHEW. Senator, I cannot talk about whether there's any ongoing——

Senator COTTON. Are you currently being sued by the Federal Trade Commission?

Mr. CHEW. Senator, I cannot talk about any potential lawsuits, whether they happen——

Senator COTTON. I didn't say potential—actual. Are you being sued by the Federal Trade Commission?

Mr. CHEW. Senator, I think I've given you my answer. I cannot talk about——

Senator COTTON. The answer's no. Ms. Yaccarino's company is being sued, I believe. Mr. Zuckerberg's company is being sued, I believe. Yet, TikTok, the agent of the Chinese Communist Party is not being sued by the Biden administration. Are you familiar with the name Cristina Caffarra?

Mr. CHEW. You may have to give me more details.

Senator COTTON. Cristina Caffarra was a paid advisor to ByteDance, your Communist-influenced parent company. She was then hired by the Biden FTC to advise on how to sue Mr. Zuckerberg's company.

Mr. CHEW. Senator, ByteDance is a global company and not a Chinese Communist company. It's owned by global investors.

Senator COTTON. Public reports indicate that your lobbyist visited the White House more than 40 times in 2022. How many times did your company's lobbyist visit the White House last year?

Mr. CHEW. I don't know that, Senator.

Senator COTTON. Are you aware that the Biden campaign and the Democratic National Committee is on your platform, they have TikTok accounts?

Mr. CHEW. Senator, we encourage people to come on——

Senator COTTON. Which, by the way——

Mr. CHEW [continuing]. To create content.

Senator COTTON [continuing]. They won't let their staffers use their personal phones. They give them separate phones that they only use TikTok on.

Mr. CHEW. We encourage everyone to join, including yourself, Senator.

Senator COTTON. So all these companies are being sued by the FTC. You're not. The FTC has a former paid advisor of your parent talking about how they can sue Mr. Zuckerberg's company. Joe Biden's reelection campaign, the Democratic National Committee is on your platform. Let me ask you, have you or anyone else at TikTok communicated with or coordinated with the Biden administration, the Biden campaign, or the Democratic National Committee to influence the flow of information on your platform?

Mr. CHEW. We work with anyone, any creators who want to use our campaign. It's all the same process that we have——

Senator COTTON. Okay. So what we have here, we have a company that's a tool of the Chinese Communist Party that is poisoning the minds of America's children, in some cases, driving them to suicide. And that at best, the Biden administration is taking a pass on, at worse, may be in collaboration with. Thank you, Mr. Chew.

Chair DURBIN. Thank you, Senator Cotton. So we're going to take a break now. We're on the second roll call. Members can take advantage of if they wish. The break will last about 10 minutes. Please do your best to return.

[Whereupon the hearing was recessed and reconvened.]

Chair DURBIN. The Senate Judiciary Committee will resume. We have nine Senators who have not asked questions yet, in 7-minute rounds, and we'll turn first to Senator Padilla.

Senator PADILLA. Thank you, Mr. Chair. Colleagues, as we reconvene, I'm proud once again to share that I am one of the few Senators with younger children. And I lead with that because as we are having this conversation today, it's not lost on me that between my children, who are all now in a teen and preteen category, and their friends, I see this issue very up close and personal.

And in that spirit, I want to take a second to just acknowledge and thank all the parents who are in the audience today, many of

whom have shared their stories with our offices. And I credit them for finding strength through their suffering, through their struggle, and channeling that into the advocacy that is making a difference. I thank all of you.

Now, I appreciate, again, personally, the challenges that parents, and caretakers, school personnel, and others face in helping our young people navigate this world of social media and technology in general. Now, the services our children are growing up with provide them unrivaled access to information. I mean, this is beyond what previous generations have experienced, and that includes learning opportunities, socialization, and much, much more.

But we also clearly have a lot of work to do to better protect our children from the predators and predatory behavior that these technologies have enabled. And yes, Mr. Zuckerberg, that includes exacerbating the mental health crisis in America.

Nearly all teens we know have access to smartphones and the internet and use the internet daily. And while guardians do have primary responsibility for caring for our children, the old adage says, "it takes a village," and so society as a whole, including leaders in the tech industry, must prioritize the health and safety of our children.

Now, I'll dive into my questions now and be specific, platform by platform, witness by witness on the topic of some of the parental tools you have each made reference to.

Mr. Citron, how many minors are on Discord, and how many of them have caretakers that have adopted your Family Center tool? And if you don't have the numbers, just say that quickly and provide that to our office.

Mr. CITRON. We can follow up with you on that.

Senator PADILLA. How have you ensured that young people and their guardians are aware of the tools that you offer?

Mr. CITRON. We make it very clear to use it—to teens on our platform what tools are available——

Senator PADILLA. That sounds very vague.

Mr. CITRON [continuing]. And our Teen Safety Assist is enabled by default,

Senator PADILLA. What specifically do you do? What may be clear to you is not clear to the general public. So what do you do, in your opinion, to make it very clear?

Mr. CITRON. So our Teen Safety Assist, which is a feature that helps teens keep themselves safe in addition to blocking and blurring images that may be sent to them, that is on by default for teen accounts, and it cannot be turned off. We market to our teen users directly on our platform, we launched our Family Center. We create a promotional video, and we put it directly on our product. So when every teen opened the app, in fact, every user opened the app, they got an alert like, Hey, Discord has this. They want you to use it.

Senator PADILLA. Thank you. Look forward to the data that we're requesting.

Mr. Zuckerberg, across all of Meta services from Instagram, Facebook, Messenger, and Horizon, how many minors use your applications? And of those minors, how many have a caretaker that has adopted the parental supervision tools that you offer?

Mr. ZUCKERBERG. I can follow up with the specific stats on that, Senator.

Senator PADILLA. Okay. It would be very helpful not just for us to know, but for you to know as a leader of your company. Same question, how are you ensuring that young people and their guardians are aware of the tools that you offer?

Mr. ZUCKERBERG. We run pretty extensive ad campaigns both on our platforms and outside. We work with creators and organizations like Girl Scouts to make sure that there's broad awareness of the tools.

Senator PADILLA. Okay. Mr. Spiegel, how many minors use Snapchat, and of those minors, how many have caretakers that are registered with your Family Center?

Mr. SPIEGEL. I believe in the United States, there are approximately 20 million teenage users of Snapchat. I believe approximately 200,000 parents use Family Center, and about 400,000 teens have linked their account to their parents using Family Center.

Senator PADILLA. So 200,000 and 400,000. Sounds like a big number, but small in percentage of the minors using Snapchat. What are you doing to ensure that young people and their guardians are aware of the tools you offer?

Mr. SPIEGEL. Senator, we create a banner for Family Center on the user's profiles so that accounts we believe maybe of the age, that they could be parents, can see the entry point into Family Center easily.

Senator PADILLA. Okay. Mr. Chew, how many minors are on TikTok, and how many of them have a caregiver that uses your family tools?

Mr. CHEW. Senator, I need to get back to you on the specific numbers. But we were one of the first platforms to give what we call family pairing to parents. You go to settings, you turn on the QR code—your teenager's QR code, and yours—you scan it. And what it allows you to do is you can set screen time limits, you can filter out some keywords, you can turn on a more restricted mode. And we are always talking to parents. I met, you know, a group of parents, and teenagers, and high school teachers last week to talk about what more we can provide in the family pairing mode.

Senator PADILLA. Ms. Yaccarino, how many minors use X, and are you planning to implement safety measures or guidance for caretakers like your peer companies have?

Ms. YACCARINO. Thank you, Senator. Less than 1 percent of all U.S. users are between the ages of 13 and 17.

Senator PADILLA. Less than 1 percent of how many?

Ms. YACCARINO. Of 90 million U.S. users.

Senator PADILLA. Okay. So still hundreds of thousands continue?

Ms. YACCARINO. Yes, yes, and every single one is very important. Being a 14-month-old company, we have reprioritized child protection and safety measures, and we have just begun to talk about and discuss how we can enhance those with parental controls.

Senator PADILLA. Let me continue with the follow-up question for Mr. Citron. In addition to keeping parents informed about the nature of various internet services, there's a lot more we obviously need to do.

For today's purposes, while many companies offer a broad range of quote unquote user empowerment tools, it's helpful to understand whether young people even find these tools helpful. So I appreciate you sharing your Teen Safety Assist, and the tools, and how you're advertising it, but have you conducted any assessments of how these features are impacting minor's use of your platform?

Mr. CITRON. Our intention is to give teens tools, capabilities, that they can use to keep themselves safe, and also, so our teams can help keep teens safe. We recently launched Teen Safety Assist last year, and I do not have a study off the top of my head, but we'd be happy to follow up with you on that.

Senator PADILLA. Okay. My time is up. I'll have follow-up questions for each of you, either in the second round or through statements for the record on a similar assessment of the tools that you've proposed.

Senator PADILLA. Thank you, Mr. Chair.

Chair DURBIN. Thank you, Senator Padilla. Senator Kennedy.

Senator KENNEDY. Thank you all for being here. Mr. Spiegel, I see you hiding down there. What does yadda yadda yadda mean?

Mr. SPIEGEL. I'm not familiar with the term Senator.

Senator KENNEDY. Very uncool. Can we agree that what you do, not what you say, what you do is what you believe and everything else is just cottage cheese?

Mr. SPIEGEL. Yes, Senator.

Senator KENNEDY. Do you agree with that? Speak up. Don't be shy. I've listened to you today. I've heard a lot of yadda yadda-ying, and I've heard you talk about the reforms you've made, and I appreciate them. And I've heard you talk about the reforms you're going to make, but I don't think you're going to solve the problem. I think Congress is going to have to help you. I think the reforms you're talking about, to some extent, are going to be like putting paint on rotten wood.

And I'm not sure you're going to support this legislation. I'm not. The fact is that you and some of your internet colleagues who are not here, are no longer—you're not companies, you're countries. You're very, very powerful, and you and some of your colleagues who are not here have blocked everything we have tried to do in terms of reasonable regulation. Everything from privacy to child exploitation.

And in fact, we have a new definition of recession. A recession is when—we know we're in a recession when Google has to lay off 25 Members of Congress. That's what we're down to. We're also down to this fact: that your platforms are hurting children. I'm not saying they're not doing some good things, but they're hurting children.

And I know how to count votes, and if this bill comes to the floor of the U.S. Senate, it will pass. What we're going to have to do—and I say this with all the respect that I can muster—is convince my good friend, Senator Schumer, to go Amazon, by spying online and bring this bill to the Senate floor, and the House will then pass it. Now, that's one person's opinion. I may be wrong, but I doubt it.

Mr. Zuckerberg, let me ask you a couple of questions. Might wax a little philosophical here. I have to hand it to you. You have con-

vinced over 2 billion people to give up all of their personal information, every bit of it, in exchange for getting to see what their high school friends had for dinner Saturday night. That's pretty much your business model, isn't it?

Mr. ZUCKERBERG. It's not how I would characterize it. And we give people the ability to connect with the people they care about, and to engage with the topics that they care about.

Senator KENNEDY. And you take this information, this abundance of personal information, and then you develop algorithms to punch people's hot buttons, and steer to them information that punches their hot buttons again, and again, and again to keep them coming back and to keep them staying longer. And as a result, your users see only one side of an issue. And so, to some extent, your platform has become a killing field for the truth, hasn't it?

Mr. ZUCKERBERG. I mean, Senator, I disagree with that characterization. You know, we build ranking and recommendations because people have a lot of friends and a lot of interests, and they want to make sure that they see the content that's relevant to them. We're trying to make a product that's useful to people, and make our services as helpful as possible for people to connect with the people they care about and the interest they care about.

Senator KENNEDY. But you don't show them both sides. You don't give them balanced information. You just keep punching their hot buttons, punching their hot buttons. You don't show them balanced information so people can discern the truth for themselves, and you rev them up so much that so often your platform and others becomes just cesspools of snark where nobody learns anything, don't they?

Mr. ZUCKERBERG. Well, Senator, I disagree with that. I think people can engage in the things that they're interested in and learn quite a bit about those. We have done a handful of different experiments and things in the past around news and trying to show content on, you know, diverse set of perspectives. I think that there's more that needs to be explored there, but I don't think that we can solve that by ourselves. One of the things that I saw——

Senator KENNEDY. Do you think—I'm sorry to cut you off, Mr. President, but I'm going to run out of time. Do you think your users really understand what they're giving to you, all their personal information, and how you process it, and how you monetize it? Do you think people really understand?

Mr. ZUCKERBERG. Senator, I think people understand the basic terms. I mean, I think that there's—I actually think that a lot of people overestimate the amount of information we have——

Senator KENNEDY. Let me put it another way. We spent a couple years since we talked about this. Does your user agreements still suck?

[Laughter.]

Mr. ZUCKERBERG. I'm not sure how to answer that, Senator. I think there's——

Senator KENNEDY. Can you still have a dead body in all that legalese where nobody can find it?

Mr. ZUCKERBERG. Senator, I'm not quite sure what you're referring to, but I think people get the basic deal of using these services.

It's a free service. You're using it to connect with the people you care about. If you share something with people, other people will be able to see your information. It's inherently—and if you're putting something out there to be shared publicly or with a private set of people, it's—you know, you're inherently putting it out there. So I think people get that basic part of how this works.

Senator KENNEDY. But Mr. Zuckerberg, you're in the foothills of creepy. You track people who aren't even Facebook users. You track your own people, your own users who are your product, even when they're not on Facebook.

I'm going to land this plane pretty quickly, Mr. Chairman. I mean, it's creepy, and I understand you make a lot of money doing it, but I just wonder if our technology is greater than our humanity. I mean, let me ask you this final question. Instagram is harmful to young people, isn't it?

Mr. ZUCKERBERG. Senator, I disagree with that. That's not what the research shows on balance. That doesn't mean that individual people don't have issues, and that there aren't things that we need to do to help provide the right tools for people. But across all of the research that we've done internally, I mean, this—you know, the survey that the Senator previously cited, you know, there are 12 or 15 different categories of harm that we asked teens if they felt that Instagram made it worse or better. And across all of them, except for the one that Senator Hawley cited, more people said that using Instagram——

Senator KENNEDY. I've got to land this plane, Mr. Zuckerberg.

Mr. ZUCKERBERG [continuing]. Contributed to issues that they faced, either positive or——

Senator KENNEDY. We just have to agree to disagree. If you believe that Instagram—I'm not saying it's intentional, but if you agree that Instagram—if you think that Instagram is not hurting millions of our young people, particularly young teens, particularly young women, you shouldn't be driving it. It is. Thanks.

Chair DURBIN. Senator Butler.

Senator BUTLER. Thank you, Mr. Chair. And thank you to our panelists who've come to have an important conversation with us. Most importantly, I want to appreciate the families who have shown up to continue to be remarkable champions of your children and your loved ones, for being here, and in particular to California families that I was able to just talk to on the break. The families of Sammy Chapman from Los Angeles and Daniel Puerta from Santa Clarita. They are here today and are doing some incredible work to not just protect the memory and legacy of their boys, but the work that they're doing is going to protect my 9-year-old. And that is indeed why we are here.

There are a couple questions that I want to ask some individuals. Let me start with a question for each of you. Mr. Citron, have you ever sat with a family and talked about their experience and what they need from your product? Yes, or no?

Mr. CITRON. Yes. I have spoken with parents about how we can build tools to help them.

Senator BUTLER. Mr. Spiegel, have you sat with families and young people to talk about your products and what they need from your product?

Mr. SPIEGEL. Yes, Senator.

Senator BUTLER. Mr. Chew?

Mr. CHEW. Yes. I just did it 2 weeks ago. Like, for example——

Senator BUTLER. I don't want to know what you did for the hearing prep, Mr. Chew. I just wanted to know if——

Mr. CHEW. No, it's an example.

Senator BUTLER [continuing]. You did anything——

Mr. CHEW. Senator, it's an example.

Senator BUTLER [continuing]. In terms of designing the product that you are creating. Mr. Zuckerberg, have you sat with parents and young people to talk about how you design product for your consumers?

Mr. ZUCKERBERG. Yes. Over the years, I've had a lot of conversations with parents——

Senator BUTLER. You know, that's interesting, Mr. Zuckerberg, because we talked about this last night, and you gave me a very different answer. I asked you this very question.

Mr. ZUCKERBERG. Well, I told you that I wasn't—that I didn't know what specific processes our company had for——

Senator BUTLER. No, Mr. Zuckerberg, you said to me that you had not.

Mr. ZUCKERBERG. I must have misspoke.

Senator BUTLER. I want to give you the room to misspeak, Mr. Zuckerberg, but I asked you this very question. I asked all of you this question and you told me a very different answer when we spoke, but I won't belabor it.

A number of you have talked about the—I'm sorry, X Ms. Yaccarino, have you talked to parents directly, young people, about designing your product?

Ms. YACCARINO. As a new leader of X, the answer is yes. I've spoken to them about the behavioral patterns because less than 1 percent of our users are in that age group, but yes, I have spoken to them.

Senator BUTLER. Thank you, ma'am. Mr. Spiegel, there are a number of parents whose children have been able to access illegal drugs on your platform. What do you say to those parents?

Mr. SPIEGEL. Well, Senator, we are devastated that we cannot——

Senator BUTLER. To the parents. What do you say to those parents, Mr. Spiegel?

Mr. SPIEGEL. I'm so sorry that we have not been able to prevent these tragedies. We work very hard to block all search terms related to drugs from our platform. We proactively look for and detect drug-related content. We remove it from our platform, preserve it as evidence, and then we refer it to law enforcement for action.

We've worked together with nonprofits and with families on education campaigns because the scale of the fentanyl epidemic is extraordinary. Over 100,000 people lost their lives last year, and we believe people need to know that one pill can kill. That campaign was viewed more than 260 million times on Snapchat. We also launched——

Senator BUTLER. Mr. Spiegel, there are two fathers in this room who lost their sons. They were 16 years old. Their children were able to get those pills from Snapchat. I know that there are statis-

tics, and I know that there are good efforts. None of those efforts are keeping our kids from getting access to those drugs on your platform.

Now, as a California company, all of you, I've talked with you about what it means to be a good neighbor, and what California families and American families should be expecting from you. You owe them more than just a set of statistics. And I look forward to you showing up on all pieces of this legislation—all of you, showing up on all pieces of legislation to keep our children safe.

Mr. Zuckerberg, I want to come back to you. I talked with you about being a parent to a young child who doesn't have a phone, you know, is not on social media at all. And one of the things that I am deeply concerned with as a parent to a young black girl, is the utilization of filters on your platform that would suggest to young girls utilizing your platform, the evidence that they are not good enough as they are.

I want to ask more specifically and refer to some unredacted court documents that revealed that your own researchers concluded that these face filters that mimic plastic surgery, negatively impact youth mental health, indeed, and well-being. Why should we believe—why should we believe that because—that you are going to do more to protect young women and young girls when it is that you give them the tools to affirm the self-hate that is spewed across your platforms? Why is it that we should believe that you are committed to doing anything more to keep our children safe?

Mr. ZUCKERBERG. Sorry, there's a lot to unpack there.

Senator BUTLER. There is a lot.

Mr. ZUCKERBERG. We give people tools to express themselves in different ways, and people use face filters and different tools to make media, and photos, and videos that are fun or interesting across a lot of the different products that are——

Senator BUTLER. Plastic surgery pins are good tools to express creativity?

Mr. ZUCKERBERG. Senator, I'm not speaking to that specifically.

Senator BUTLER. Skin lightening tools are tools to express creativity?

Mr. ZUCKERBERG. Senator, I'm not——

Senator BUTLER [continuing]. This is the direct thing that I'm asking about.

Mr. ZUCKERBERG. I'm not defending any specific one of those. I think that the ability to kind of filter and edit images is generally a useful tool for expression for that specifically. I'm not familiar with the study that you're referring to, but we did make it so that we're not recommending this type of content to teens.

Senator BUTLER. I made no reference to a study. To court documents that revealed your knowledge of the impact of these types of filters on young people, generally young girls in particular, and——

Mr. ZUCKERBERG. Senator, I disagree with that characterization. I think that——

Senator BUTLER. With court documents?

Mr. ZUCKERBERG [continuing]. There have been hypothesis—I haven't seen any documents that says——

Senator BUTLER. Okay. Mr. Zuckerberg, my time is up. I hope that you hear what is being offered to you, and are prepared to step up and do better. I know this Senate Committee is going to do our work to hold you to greater account. Thank you, Mr. Chair.

Chair DURBIN. Senator Tillis.

Senator TILLIS. Thank you, Mr. Chair. Thank you all for being here. I don't feel like I'm going to have an opportunity to ask a lot of questions, so I'm going to reserve the right to submit some for the record.

Senator TILLIS. We've had hearings like this before. I've been in the Senate for 9 years. I've heard hearings like this before. I've heard horrible stories about people who have died, committed suicide, been embarrassed. Every year we have an annual flogging, every year. And what materially has occurred over the last 9 years? Do any of you-all—just a yes or no question. Do any of you-all participate in an industry consortium trying to make this fundamentally safe across platforms? Yes, or no, Mr. Zuckerberg?

Mr. ZUCKERBERG. Yes.

Ms. YACCARINO. There's a variety of organizations that we work——

Senator TILLIS. Do you participate in them?

Ms. YACCARINO. Which organization, Senator?

Senator TILLIS. I should say, does anyone here not participate in an industry—I actually think it would be immoral for you all to consider it a strategic advantage to keep safe—or, to keep private something that would secure all these platforms to avoid this sort of problem. Do you-all agree with that? That anybody that would be saying, you want ours because ours is the safest, and these haven't figured out the secret sauce—that you as an industry realize this is an existential threat to you-all if we don't get it right. Right?

I mean, you've got to secure your platforms. You got to deal with this. Do you not have an inherent mandate to do this? Because it would seem to me if you don't, you're going to cease to exist. I mean, we could regulate you out of business if we wanted to.

And the reason I'm saying, it may sound like criticism, it's not a criticism. I think we have to understand that there should be an inherent motivation for you to get this right. Our Congress will make a decision that could potentially put you out of business.

Here's the reason I have a concern with that though. I just went on the internet while I was listening intently to all the other Members speaking, and I found a dozen different platforms outside the United States. Ten of which are in China, two of which are in Russia. Their daily average subscriber or active membership numbers in the billions. Well, people say you can't get on China's version of TikTok. I took me one quick search on my favorite search engine to find out exactly how I could get an account on this platform today.

And so the other thing that we have to keep in mind, I come from technology. I could figure out, ladies and gentlemen, I could figure out how to influence your kid without them ever being on a social media platform. I can randomly send texts and get a bite, and then find out an email address and get compromising information.

It is horrible to hear some of these stories. And I have shared the—and I've had these stories occur in my hometown down in North Carolina. But if we only come here and make a point today, and don't start focusing on making a difference, which requires people to stop shouting, and start listening, and start passing language here, the bad actors are just going to be off our shores.

I have another question for you all. How many people roughly—if you don't know the exact numbers, okay. Roughly, how many people do you have looking 24 hours a day at these horrible images, and just go real quick with an answer down the line, and filtering it out?

Mr. ZUCKERBERG. It's most of the 40,000, about, people who work on safety.

Senator TILLIS. And again.

Ms. YACCARINO. We have 2,300 people all over the world.

Senator TILLIS. Okay.

Mr. CHEW. We have 40,000 trust and safety professionals around the world.

Mr. SPIEGEL. We have approximately 2,000 people dedicated to trust and safety and content moderation.

Mr. CITRON. Our platform is much, much smaller than these folks. We have hundreds of people. And it's looking at the content, and 15 percent of our workforce is focused on it.

Senator TILLIS. I've already mentioned, these people have a horrible job. Many of them experience—they have to get counseling for all the things they see. We have evil people out there, and we're not going to fix this by shouting past or talking past each other. We're going to fix this by every one of you-all being at the table, and hopefully coming closer to what I heard one person say, supporting a lot of the good bills, like one that I hope Senator Blackburn mentions when she gets a chance to talk.

But guys, if you're not at the table and securing these platforms, you're going to be on it. And the reason why I'm not okay with that is that if we ultimately destroy your ability to create value and drive you out of business, the evil people will find another way to get to these children.

And I do have to admit—I don't think my mom's watching this one—but there is good. We can't look past good that is occurring. My mom, who lives in Nashville, Tennessee, and I talked to her yesterday and we talked about a Facebook post that she made a couple of days ago. We don't let her talk to anybody else. That connects my 92-year-old mother with her grandchildren and great-grandchildren. That lets a kid who may feel awkward in school to get into a group of people and relate to people. Let's not throw out the good because we haven't all together focused on rooting out the bad.

Now, I guarantee you, I could go through some of your governance documents and find a reason to flog every single one of you because you didn't place the emphasis on it that I think you should. But at the end of the day, I find it hard to believe that any of you people started this business, some of you in your college dorm rooms, for the purposes of creating the evil that is being perpetrated on your platforms.

But I hope that every single waking hour, you are doing everything you can to reduce it. You're not going to be able to eliminate it. And I hope that there are some enterprising young tech people out there today that are going to go to parents and say, ladies and gentlemen, your children have a deadly weapon. They have a potentially deadly weapon, whether it's a phone or a tablet. You have to secure it. You can't assume that they're going to be honest and say that they're 16 when they're 12.

We all have to recognize that we have a responsibility to play and you guys are at the tip of the spear. So I hope that we can get to a point to where we are moving these bills. If you got a problem with them, state your problem. Let's fix it. No is not an answer. And know that I want the United States to be the beacon for innovation, to be the beacon for safety, and to prevent people from using other options that have existed since the internet has existed to exploit people, and count me in as somebody that will try and help out. Thank you, Mr. Chair.

Chair DURBIN. Thank you, Senator Tillis. Next is Senator Ossoff.

Senator OSSOFF. Thank you, Mr. Chairman. And thank you to our witnesses today. Mr. Zuckerberg, I want to begin by just asking a simple question, which is, do you want kids to use your platform more or less?

Mr. ZUCKERBERG. Well, we don't want people under the age of 13 using our——

Senator OSSOFF. Do you want teenagers 13 and up to use your platform more or less?

Mr. ZUCKERBERG. Well, we would like to build a product that is useful and that people want to use it more.

Senator OSSOFF. My time is going to be limited. So do you want them to use it more or less? Teenagers, 13 to 17 years old, do you want them using Meta products more or less?

Mr. ZUCKERBERG. I'd like them to be useful enough that they want to use them more.

Senator OSSOFF. You want them to use it more. I think herein we have one of the fundamental challenges. In fact, you have a fiduciary obligation, do you not, to try to get kids to use your platform more?

Mr. ZUCKERBERG. It depends on how you define that. We obviously are a business, but——

Senator OSSOFF. I'm sorry, Mr. Zuckerberg, it's self-evident that you have a fiduciary obligation to get your users, including users under 18, to use and engage with your platform more rather than less. Correct?

Mr. ZUCKERBERG. Over the long-term. But in the near-term, we often take a lot of steps, including we made a change to show less videos on the platform. That reduced the amount of time by more than 50 million hours.

Senator OSSOFF. Okay. But if your shareholders asked you, "Mark—" I wouldn't, Mr. Zuckerberg here, but your shareholders might be on a first name basis with you. "Mark, are you trying to get kids to use Meta products more or less?" You'd say more, right?

Mr. ZUCKERBERG. Well, I would say that over the long term, we're trying to create the most value——

Senator OSSOFF. Yes. So the 10–K you filed with the SEC, a few things I want to note here are some quotes. And this is a filing that you signed, correct?

Mr. ZUCKERBERG. Yes.

Senator OSSOFF. Yes. "Our financial performance has been and will continue to be significantly determined by our success in adding, retaining, and engaging active users." Here's another quote. "If our users decrease their level of engagement with our products, our revenue, financial results, and business may be significantly harmed."

Here's another quote. "We believe that some users, particularly younger users, are aware of and actively engaging with other products and services, similar to, as a substitute for ours." It continues, "In the event that users increasingly engage with other products and services, we may experience a decline in use and engagement in key demographics or more broadly, in which case our business would likely be harmed."

You have an obligation as the chief executive to encourage your team to get kids to use your platform more.

Mr. ZUCKERBERG. Senator, I think this is——

Senator OSSOFF. Is that not self-evident? You have a fiduciary obligation to your shareholders to get kids to use your platform more.

Mr. ZUCKERBERG. I think that the thing that's not intuitive is the direction is to make the products more useful so that way people want to use them more. We don't give the teams running the Instagram feed or the Facebook feed a goal to increase the amount of time that people spend.

Senator OSSOFF. Yes. But you don't dispute and your 10–K makes it clear you want your users engaging more and using more the platform. And I think this gets to the root of the challenge because it's the overwhelming view of the public. Certainly, in my home State of Georgia.

And we've had some discussions about the underlying science that this platform is harmful for children. I mean, you are familiar with, and not just your platform, by the way, social media in general, 2023 report from the Surgeon General about the impact of social media on kids' mental health, which cited evidence that kids who spend more than 3 hours a day on social media have double the risk of poor mental health outcomes, including depression and anxiety. Are you familiar with that Surgeon General report and the underlying study?

Mr. ZUCKERBERG. I read the report. Yes.

Senator OSSOFF. Do you dispute it?

Mr. ZUCKERBERG. No, but I think it's important to characterize it correctly. I think what he was flagging in the report is that there seems to be a correlation, and obviously the mental health issue is very important. So it's something that needs to be studied further.

Senator OSSOFF. Yes, everyone knows there's a correlation. Everyone knows that kids who spend a lot of time, too much time on your platforms are at risk. And it's not just the mental health issues—let me ask you another question. Is your platform safe for kids?

Mr. ZUCKERBERG. I believe it is, but there's a——

Senator OSSOFF. Hold on a second. Let me ask you——

Mr. ZUCKERBERG [continuing]. Difference between correlation and causation.

Senator OSSOFF [continuing]. Because we're not going to be able to get anywhere. We want to work in a productive, open, honest, and collaborative way with the private sector to pass legislation that will protect Americans, that will protect American children above all, and that will allow businesses to thrive in this country. If we don't start with an open, honest, candid, realistic assessment of the issues, we can't do that.

The first point is you want kids to use the platform more. In fact, you have an obligation to. But if you're not willing to acknowledge that it's a dangerous place for children—the internet is a dangerous place for children, not just your platform, isn't it? Isn't the internet a dangerous place for children?

Mr. ZUCKERBERG. I think it can be. Yes. There's both great things that people can do and there are harms that we need to work toward—yes.

Senator OSSOFF. It's a dangerous place for children. There are families here who have lost their children. There are families across the country whose children have engaged in self-harm, who have experienced low self-esteem, who have been sold deadly pills on the internet. The internet's a dangerous place for children, and your platforms are dangerous places for children. Do you agree?

Mr. ZUCKERBERG. I think that there are harms that we need to work to mitigate. I mean, I'm not going to——

Senator OSSOFF. Why not? Why not just acknowledge it? Why do we have to do the very careful code?

Mr. ZUCKERBERG. Well, I just disagree with the characterization——

Senator OSSOFF. Which characterization? That the internet's a dangerous place for children?

Mr. ZUCKERBERG. I think you're trying to characterize our products as inherently dangerous, and I think that's——

Senator OSSOFF. Inherent or not, your products are places where children can experience harm. They can experience harm to their mental health. They can be sold drugs. They can be preyed upon by predators. They're dangerous places, and yet you have an obligation to promote the use of these platforms by children.

And look, all I'm trying to suggest to you, Mr. Zuckerberg, and my time is running short, is that in order for you to succeed, you and your colleagues here, we have to acknowledge these basic truths. We have to be able to come before the American people, the American public, the people in my State of Georgia, and acknowledge the internet is dangerous, including your platforms. There are predators lurking. There are drugs being sold. There are harms to mental health that are taking a huge toll on kids' quality of life.

And yet you have this incentive, not just you, Mr. Zuckerberg, all of you have an incentive to boost, maximize use, utilization, and engagement. And that is where public policy has to step in to make sure that these platforms are safe for kids so kids are not dying, so kids are not overdosing, so kids are not cutting themselves or killing themselves because they're spending all day scrolling instead of playing outside. And I appreciate all of you for your testi-

mony. We will continue to engage as we develop this legislation. Thank you.

Chair DURBIN. Senator from Tennessee.

Senator BLACKBURN. Thank you, Mr. Chairman. Thank you to each of you for coming. And I know some of you had to be subpoenaed to get here, but we do appreciate that you-all are here.

Mr. Chew, I want to come to you first. We've heard that you're looking at putting a headquarters in Nashville, and likewise in Silicon Valley and Seattle. And what you're going to find probably is that the welcome mat is not going to be rolled out for you in Nashville like it would be in California. There are a lot of people in Tennessee that are very concerned about the way TikTok is basically building dossiers on our kids, the way they are building those on their Virtual U. And also, that that information is held in China, in Beijing, as you responded to Senator Blumenthal and I last year in reference to that question.

And we also know that a major music label yesterday said they were pulling all of their content off your site because of your issues on payment, on artificial intelligence, and because of the negative impact on our kids' mental health. So we will see how that progresses.

Mr. Zuckerberg, I want to come to you. We have just had, Senator Blumenthal and I, of course, have had some internal documents in emails that have come our way. One of the things that really concerned me is that you referred to your young users in terms of their lifetime value of being roughly $270 per teenager. And each of you should be looking at these kids, their T-shirts they're wearing today say, "I'm worth more than $270." We've got some standing up in those t-shirts.

[Applause.]

Senator BLACKBURN. And some of the children from our State, some of the children, the parents that we have worked with, just to think whether it is Becca Schmidt, David Malloch, Sarah Flatt, and Lee Schopt, would you say that life is only worth $270? What could possibly lead you—I mean, I listened to that. I know you're a dad, I'm a mom, I'm a grandmom. And how could you possibly even have that thought? It's astounding to me.

And I think this is one of the reasons that States, 42 States, are now suing you because of features that they consider to be addictive, that you are pushing forward. And in the emails that we've got from 2021, that go from August to November, there is the Staff Plan that is being discussed. And Antigone Davis, Nick Clegg, Cheryl Sandberg, Chris Cox, Alex Schultz, Adam Mosseri, are all on this chain of emails on the well-being plan. And then we get to one, "Nick did email Mark for—to emphasize his support for the package, but it sounds like it lost out to various other pressures and priorities."

See, this is what bothers us. Children are not your priority. Children are your product. Children you see as a way to make money, and protecting children in this virtual space—you made a conscious decision even though Nick Clegg and others were going through the process of saying this is what we do. These documents are really illuminating. And it just shows me that growing this business, expanding your revenue, what you were going to put on those quar-

terly filings, that was the priority. The children were not. It's very clear.

I want to talk with you about the pedophile ring because that came up earlier and The Wall Street Journal reported on that. And one of the things that we found out was after that became evident, then you didn't take that content down. And it was content that showed that teens were for sale and were offering themselves to older men.

And you didn't take it down because it didn't violate your community standards. Do you know how often a child is bought or sold for sex in this country? Every 2 minutes. Every 2 minutes a child is bought or sold for sex. That's not my stat. That is a TBI stat.

Now finally this content was taken down after a congressional staffer went to Meta's global head of safety. So would you please explain to me and to all these parents why explicit predatory content does not violate your platform's terms of service or your community standards?

Mr. ZUCKERBERG. Sure, Senator, let me try to address all of the things that you just said. It does violate our standards. We work very hard to take it down.

Senator BLACKBURN. Didn't take it down.

Mr. ZUCKERBERG. Well, we've reported, I think it's more than 26 million examples of this kind of content.

Senator BLACKBURN. Didn't take it down until a congressional staffer brought it up.

Mr. ZUCKERBERG. It may be that in this case we made a mistake and missed something. But we have——

Senator BLACKBURN. I think you make a lot of mistakes——

Mr. ZUCKERBERG [continuing]. Leading teams that identify more than——

Senator BLACKBURN [continuing]. So let's move. I want to talk with you about your Instagram creators program, and about the push we found out through these documents that you actually are pushing forward because you want to bring kids in early. You see these younger tweenagers as, "valuable, but an untapped audience," quoting from the emails, and suggesting teens are actually household influencers to bring their younger siblings into your platform, into Instagram.

Now, how can you ensure that Instagram creators, your product, your program, does not facilitate illegal activities when you fail to remove content pertaining to the sale of minors. And it's happening once every 2 minutes in this country.

Mr. ZUCKERBERG. Senator, our tools for identifying that kind of content are industry-leading. That doesn't mean we're perfect. There are definitely issues that we have, but we continue to invest——

Senator BLACKBURN. Mr. Zuckerberg, yes, there are a lot that is slipping through. It appears that you're trying to be the premier sex trafficking site.

Mr. ZUCKERBERG. Of course not, Senator.

Senator BLACKBURN [continuing]. In this country.

Mr. ZUCKERBERG. Senator, that's ridiculous.

Senator BLACKBURN. It's not ridiculous. You want to turn around and tell the people that——

Mr. ZUCKERBERG. We don't want this content on our platforms, and we——

Senator BLACKBURN. Why don't you take it down?

Mr. ZUCKERBERG. We do take it down?

Senator BLACKBURN. We're here discussing——

Mr. ZUCKERBERG. We do more work——

Senator BLACKBURN. We need you to work with us——

Mr. ZUCKERBERG [continuing]. To take it down than——

Senator BLACKBURN. No, you are not. You are not. And the problem is, we've been working on this—Senator Welch is over there. We've been working on this stuff for a decade. You have an army of lawyers and lobbyists that have fought us on this every step of the way. You work with Net Choice, the Cato Institute, Taxpayers Protection Alliance, and Chamber of Progress to actually fight our bipartisan legislation to keep kids safe online. So, are you going to stop funding these groups? Are you going to stop lobbying against this, and come to the table and work with us? Yes, or no?

[Applause.]

Mr. ZUCKERBERG. Senator, we have a——

Senator BLACKBURN. Yes or no?

Mr. ZUCKERBERG. Of course, we'll work with you on the legislation. I mean, it's——

Senator BLACKBURN. Okay. The door is open. We've got all these bills, you need to come to the table. Each and every one of you need to come to the table, and you need to work with us. Kids are dying.

[Applause.]

Chair DURBIN. Senator Welch.

Senator WELCH. I want to thank my colleague, Senator Blackburn for her decade of work on this. I actually have some optimism. There is a consensus today that didn't exist, say, 10 years ago, that there is a profound threat to children, to mental health, to safety. There's not a dispute; that was in debate before. That's a starting point.

Secondly, we're identifying concrete things that can be done in four different areas. One is industry standards, two is legislation, three are the courts, and then four, is a proposal that Senator Bennett, Senator Graham, myself, and Senator Warren have to establish an agency, a Governmental agency whose responsibility would be to engage in this on a systematic, regular basis with proper resources.

And I just want to go through those. I appreciate the industry standard decisions and steps that you've taken in your companies, but it's not enough. And that's what I think you're hearing from my colleagues. Like for instance, where there are layoffs is in the trust and verify programs. That's alarming because it looks like there is a reduction in emphasis on protecting things. Like you just added, Ms. Yaccarino, 100 employees in Texas in this category. And how many did you have before?

Ms. YACCARINO. The company is just coming through a significant restructuring. So we've increased the number of trust and safety employees and agents all over the world by at least 10 percent so far in the last 14 months, and we'll continue to do so specifically in Austin, Texas.

Senator WELCH. All right. Mr. Zuckerberg, my understanding is there have been layoffs in that area as well. There's added jobs there at Twitter, but at Meta, have there been reductions in that?

Mr. ZUCKERBERG. There have been across the board, not really focused on that area. I think our investment is relatively consistent over the last couple of years. We invested almost $5 billion in this work last year, and I think this year, we'll be on the same order of magnitude.

Senator WELCH. All right. And another question that's come up is when to the horror of a user of any of your platforms, somebody has an image on there that's very compromising, often of a sexual nature, is there any reason in the world why a person who wants to take that down can't have a very simple same day response to have it taken down? I'll start with Twitter.

Ms. YACCARINO. I'm sorry, Senator. I was taking notes. Could you repeat the question?

Senator WELCH. Well, there's a lot of examples of a young person of finding out about an image that is of them, and really compromises them, and actually can create suicidal thoughts and they want to call up or they want to send an email and say, take it down. I mean, why is it not possible for that to be responded to immediately?

Ms. YACCARINO. Well, we all strive to take down any type of violative content or disturbing content immediately. At X, we have increased our capabilities with a two-step reporting process.

Senator WELCH. Shouldn't it just be standard? If I'm a parent or I'm a kid and I want this down, shouldn't there be methods in place where it comes down? You can see what the image is.

Ms. YACCARINO. And an ecosystem-wide standard would improve and actually enhance the experience for users at all our platforms.

Mr. ZUCKERBERG. There actually is an organization that I think a number of the companies up here are a part of called Take It Down. It's some technology that we and a few others built, but basically——

Senator WELCH. So you-all are in favor of that because——

Mr. ZUCKERBERG. Oh yes, this already exists.

Senator WELCH [continuing]. Then it's going to give some peace of mind to people. All right? It really, really matters. I don't have that much time. So we've talked about the legislation, and Senator Whitehouse had asked you to get back with your position on Section 230, which I'll go to in a minute. But I would welcome each of you responding as to your company's position on the bills that are under consideration in this hearing. All right? I'm just asking you to do that.

Third, the court. This big question of Section 230. And today, I'm pretty inspired by the presence of the parents who have turned their extraordinary grief into action and hope that other parents may not have to suffer what for them is a devastating—for everyone, a devastating loss.

Senator Whitehouse asked you all to get back very concretely about Section 230 and your position on that. But it's an astonishing benefit that your industry has that no other industry has. They just don't have to worry about being held accountable in court if they're negligent. So you've got some explaining to do, and I'm just

reinforcing Senator Whitehouse's request that you get back specifically about that.

And then finally, I want to ask about this notion. It's this idea of a Federal agency who's resourced and whose job is to be dealing with public interest matters that are really affected by Big Tech. It's extraordinary what has happened in our economy with technology and your companies represent innovation and success.

But just as when the railroads were ascendant, and were in charge and ripping off farmers because of practices they were able to get away with; just as when Wall Street was flying high, but there was no one regulating blue sky laws, we now have a whole new world in the economy. And Mr. Zuckerberg, I remember you testifying in the Energy and Commerce Committee, and I asked you your position on the concept of a Federal regulatory agency. My recollection is that you were positive about that. Is that still the case?

Mr. ZUCKERBERG. I think it could be a reasonable solution. There are obviously pros and cons to doing that versus through the normal—the current structure of having different regulatory agencies focused on specific issues. But because a lot of the things tradeoff against each other, like one of the topics that we talked about today is encryption, and that's obviously really important for privacy and security.

Senator WELCH. Can we just go down the line? I'm at the end, but thank you. Ms. Yaccarino.

Ms. YACCARINO. Senator, I think the industry initiative to keep those conversations going would be something X would be very, very proactive about. If you think about our support of the REPORT Act, the SHIELD Act, the STOP CSAM Act, our support of the Project Safe Child Act, I think our intentions are clear to participate in lead here.

Senator WELCH. Mr. Chew.

Mr. CHEW. Senator, we support national privacy legislation, for example. So that sounds like a good idea. We just need to understand what it means.

Senator WELCH. All right. Mr. Spiegel.

Mr. SPIEGEL. Senator, we'll continue to work with your team, and we'd certainly be open to exploring the right regulatory body for big technology.

Senator WELCH. But the idea of a regulatory body is something that you can see has merit?

Mr. SPIEGEL. Yes, Senator.

Senator WELCH. And Mr. Citron.

Mr. CITRON. Yes. We're very open to working with you, and our peers, and anybody, on helping make the internet a safer place. You know, I think you mentioned this is not a one platform problem, right? So we do look to collaborate with other companies, and with nonprofits, and the Government.

Senator WELCH. Okay. I thank you all. Mr. Chairman, I yield back.

Chair DURBIN. Thank you, Senator Welch. Well, we're going to conclude this hearing, and thank you all for coming today. You probably have your scorecard out there. You've met, at least, 20

Members of this Committee, and have your own impressions of their questioning or approach and the like.

But the one thing I want to make clear as Chairman of this Committee for the last 3 years is this was an extraordinary vote on an extraordinary issue. A year ago, we passed five bills unanimously in this Committee. You heard all the Senators, every spot on the political spectrum was covered. Every single Senator voted unanimously in favor of the five pieces of legislation we've discussed today. It ought to tell everyone who follows Capitol Hill and Washington, a pretty stark message.

We get it, and we live it as parents and grandparents. We know what our daughters, and sons, and others are going through. They cannot cope. They cannot handle this issue on their own. They're counting on us as much as they're counting on the industry to do the responsible thing.

And some believe with impressions of our witnesses and the companies they represent, that's you're right as an American citizen, but you ought to also leave with the determination to keep the spotlight on us to do something. Not just to hold a hearing, bring out a good, strong crowd of supporters for change, but to get something done. No excuses, no excuses. We've got to bring this to a vote.

What I found in my time in the House, in the Senate, is that's the day, that's the moment of reckoning. Speeches notwithstanding, press releases, and the like. The moment of reckoning is when we call a vote on these measures. It's time to do that. I don't believe there's ever been a moment in America's wonderful history when a business or industry has stepped up and said, "Regulate us. Put some legal limits on us."

Businesses exist by and large to be profitable. And I think that we got to get behind that and say profitability at what cost. Senator Kennedy, our Republican colleague said, "Is our technology greater than our humanity?" I think that is a fundamental question that he asked. What I would add to it, are our politics greater than technology? We're going to find out. I want to thank a few people before we close up here. I've got several staffers who've worked so hard on this. Alexandra Gelber. Thank you very much, Alexandra. Jeff Hanson, Scott Robinson.

[Applause.]

Chair DURBIN. Last point I'll make, Mr. Zuckerberg, is just a little advice to you. I think your opening statement on mental health needs to be explained because I don't think it makes any sense.

There isn't a parent in this room who's had a child that's gone through an emotional experience like this that wouldn't tell you and me, "They changed right in front of my eyes. They changed. They hold themselves up in their room. They no longer reached out to their friends. They lost all interest in school."

These are mental health consequences that I think come with the abuse of this right to have access to this type of technology. So I will just—I see my colleague is—do you want to say a word?

Senator GRAHAM. I think it was a good hearing. I hope something positive comes from it. Thank you all for coming.

Chair DURBIN. The hearing record is going to remain open for a week for statements, and questions may be submitted by Senators

by 5 p.m. on Wednesday. Once again, thanks to the witnesses for coming. The hearing stands adjourned.

[Whereupon, at 1:49 p.m., the hearing was adjourned.]

[Additional material submitted for the record follows.]

A P P E N D I X

Submitted by Chair Durbin:

ADL—Fighting Hate for Good 741
AEHT—Alliance to End Human Trafficking 744
CAIDP—Center for AI and Digital Policy 769
CDT—Center for Democracy & Technology 776
End OSEAC—Coalition to Protect Kids Online 780
End OSEAC—Coalition to Protect Kids Online—
Testimony by Christine Almadjian 784
Global Survivor Network—GSN 788
Home Office—Homeland Security 791
Letter to Congress–Abigail (Survivor) 792
Letter to Congress–D.J. (Survivor) 793
Letter to Congress–Elle (Survivor) 796
Letter to Congress–Leah (Survivor) 799
Letter to Congress–Lexie (Survivor) 800
Letter to Congress–Millie (Survivor) 802
Letter to Congress–James (Parent) 803
Letter to Congress–Julia (Parent) 804
Letter to Congress–Zack (Loving Brother) 805
Letter to Congress (Redact) 807
ParentsTogether—Hart Research Associates—Polling Data Research 811
Submission for the Record—Sydnie Collins 846
Submission for the Record—Alix Fraser 849
Submission for the Record—Arielle Geismar 851
Submission for the Record—Trisha Prabhu 853
Submission for the Record—Mary Rodee 855
Submission for the Record—Uldouz Wallace 857
Survivor Parents 858
UK Approach on E2EE 862
UK Online Safety Act 864

Dangerous by Design—Council for Responsible Social Media—CRSM
https://www.govinfo.gov/content/pkg/CHRG-118shrg57444/CHRG-118shrg57444-add1.PDF

Submitted by Ranking Member Graham:

UK Approach on E2EE 862
UK Online Safety Act 864

Submitted by Senator Klobuchar:

Survivor Parents 858

Submitted by Senator Blumenthal:

Count on Mothers—KOSA—Report Findings 874
Docket 518—Zuckerberg—Motion to Dismiss 884
Docket 518–1—Zuckerberg—Motion to Dismiss—Appendix A 904
Docket 538—Zuckerberg—Opposition—Motion to Dismiss 906
Docket 555—Zuckerberg—Reply in support of—Motion to Dismiss 932
Survivor Parents 858

Submitted by Senator Blackburn:

Social Media Victims Law Center—SMVLC—Advertising Directed to Underage Kids 953

Testimony Before the US Senate Committee on the Judiciary

Written Statement of Shou Chew

Chief Executive Officer, TikTok Inc.

January 31, 2024

Chair Durbin, Ranking Member Graham, and Members of the Committee:

Thank you for the opportunity to appear before you today to talk about TikTok's commitment to the safety of our community as we seek to inspire creativity and bring joy to more than 1 billion people worldwide. In particular, I welcome the chance to address TikTok's ongoing efforts to foster a safe and age-appropriate experience for teens on the app. This is my second time appearing before Congress in the past year, and I am grateful to be able to continue the dialogue about ways in which digital platforms and lawmakers can work together to tackle important issues.

My name is Shou Chew, and I am the Chief Executive Officer of TikTok. As a father of three, the issues we will be discussing today are deeply personal to me. I firmly believe that our industry's most fundamental responsibility is to provide a safe and secure online space for our community.

TikTok's diverse community includes teenagers, centenarians, and everyone in between. Each month in the United States more than 170 million people are on TikTok, with the average member of the community being an adult over 30 years old. They come to TikTok to discover, create, and connect. None of this can happen without a sustained commitment to platform safety that is constantly evolving to address new challenges.

Last year, I made specific commitments to Congress and to the TikTok community, and we've upheld those commitments. The first commitment is: "We will keep safety—particularly for teenagers—a top priority for us." Given the focus of today's hearing, I want to expand on what that commitment means to me. It means that at TikTok, we will continue to:

1. Take a zero-tolerance approach to child sexual abuse material (CSAM) by investing in our teams and technology to help ensure we swiftly detect, remove, and report this content;
2. Make thoughtful product design choices that help make our app inhospitable to those seeking to harm teens; and
3. Partner with leading experts and other industry participants to constantly innovate and work collaboratively to solve industrywide issues.

About TikTok

TikTok is a global video-sharing entertainment platform that has empowered millions of Americans to express themselves creatively and authentically. It can only work when members of the TikTok community trust that they and their data are safe and secure.

TikTok Inc. is a US company incorporated in the United States and subject to the laws of the United States. The TikTok business is led by an executive team in the United States and Singapore and has global offices, including in Los Angeles, Silicon Valley, Nashville, New York, Washington, D.C., Dublin, London, Paris, Berlin, Dubai, Singapore, Jakarta, Seoul, and Tokyo. TikTok's headquarters are in Los Angeles and Singapore. The TikTok platform is not available in mainland China.

The ultimate parent company of TikTok Inc. is ByteDance Ltd., a privately-held, global holding company. ByteDance Ltd. is majority owned by investors around the world, and the rest of the shares are owned by the founding team and employees around the world. ByteDance Ltd.'s Board of Directors is comprised of five individuals, three of whom are American.

During my prior congressional testimony, I described the unprecedented efforts that TikTok Inc.'s US subsidiary, US Data Security Inc. (USDS), is undertaking to build a secure environment for protected TikTok US user data, protect the platform from outside influence, and implement safeguards on our content recommendation and moderation tools. Our commitment to these principles is ongoing and unwavering. Since January of 2023, new protected US TikTok user data has been stored in the Oracle Cloud in an environment controlled by USDS. Only USDS personnel are able to access protected US TikTok user data in the Oracle environment, unless authorization is given by USDS pursuant to limited exceptions, such as for legal and compliance purposes. Traffic from the Oracle Cloud now goes through Oracle controlled gateways to prevent protected US TikTok user data from being transferred to or accessed by employees of TikTok or ByteDance. The US recommendation algorithm is stored in the Oracle Cloud, trained on US user data by USDS personnel, and changes to the algorithm have to be reviewed and deployed by USDS.

TikTok continues to delete historical protected US TikTok user data from our global data centers, helping to ensure that protected US TikTok user data is stored in Oracle's Cloud and controlled by USDS. Since I testified before Congress last spring, our team has completed the first round of deletion from TikTok servers in our global data centers. In order to provide additional assurance and validation of our team's work, we have also begun independent validation efforts to confirm the comprehensiveness of the data deletions.

In the near future, we plan to announce the selection of another independent security inspector who, in addition to Oracle, will have access to our source code and systems, testing the security and integrity of TikTok's platform and its software. This is another unprecedented level of testing and security review.

We have opened a Dedicated Transparency Center (DTC) in Maryland where Oracle has access to review TikTok source code. As of this month, we have opened an additional DTC in

Colorado, and we expect to have two fully operational DTCs in the United Kingdom and Australia soon. There are independent experts reviewing TikTok code to find and eliminate security vulnerabilities.

TikTok's commitment continues to be for protected US TikTok user data to be stored in America, hosted by an American headquartered company, with access to protected US TikTok user data overseen by USDS personnel.

TikTok's Zero-Tolerance Approach to Minor Sexual Exploitation

Keeping young people safe on our platform is our priority. This starts by placing users under 13 in a separate experience designed with age-appropriate safeguards. For users who indicate they are over 13, they are placed in our 13+ experience. But if we become aware that someone in the 13+ experience is below age 13, our policy is to ban that account.

Like other platforms with user-generated content, TikTok has implemented Terms of Service and Community Guidelines designed to prioritize user safety—whether physical, psychological, social, or financial—and privacy. Our Community Guidelines establish a set of norms and a common code of conduct that help us maintain a safe and inclusive environment for our community, where genuine interactions and authentic content are encouraged. They also reflect our rigorous approach to matters of minor safety.

TikTok has eight guiding community principles that embody our commitment to platform safety. TikTok's principles are centered on balancing expression with harm prevention, embracing human dignity, and ensuring our actions are fair. They shape TikTok's day-to-day work and guide how we approach difficult enforcement decisions. TikTok's content moderation principles and practices are informed by the United Nations' Guiding Principles on Business and Human Rights and the Santa Clara Principles, and TikTok seeks to align with international legal frameworks, such as the International Bill of Human Rights and the Convention on the Rights of the Child. The very first of these community principles is to "prevent harm."

Following the enumeration of community principles, our Community Guidelines have a number of pillars, the first of which relates to Youth Safety and Well-Being. As reflected in the Community Guidelines, youth safety is a priority. ***We do not allow content that may put young people at risk of exploitation, or psychological, physical, or developmental harm.*** If we become aware of youth exploitation on our platform, we will ban the account, as well as any other accounts belonging to the person.

Our Community Guidelines also address youth safety under our Safety and Civility pillar, which makes clear that we do not allow youth exploitation and abuse, including child sexual abuse material (CSAM), nudity, grooming, sextortion, solicitation, pedophilia, and physical or psychological abuse of young people. This includes content that is real, fictional, digitally created, and shown in fine art or objects.

We report incidents of youth sexual exploitation and abuse to the National Center for Missing and Exploited Children (NCMEC). Additionally, we work with NCMEC's Take It Down service to help remove and stop the online sharing of nude, partially nude, or sexually explicit images or

videos of people under 18. And we report to relevant law enforcement authorities when we identify a specific, credible, and imminent threat to a young person's life or of serious physical injury.

TikTok's Partnerships

TikTok is continuously working to provide a safe app experience for our community, and we aim to be a leader in this area. We recognize, however, that technology is ever-evolving and that we need to be prepared to address unexpected trends and challenges as they arise. Beyond our efforts with NCMEC, TikTok works with a variety of global partners on minor safety efforts:

- We are an active member of the Tech Coalition, a global alliance of technology companies to protect children online. In addition to being on its board, we also co-led the multi-stakeholder forum on Minor Financial Sextortion held last year, and we chair committees within the organization to advance the fight against online child sexual exploitation and abuse.
- We are part of the WePROTECT Global Alliance, the largest and most diverse multi-sector alliance dedicated to ending online child sexual exploitation.
- Internet Watch Foundation (IWF) is a vital partner for TikTok in our work to counter online child sexual abuse and exploitation. TikTok accesses IWF's URL and keyword list and, via NCMEC, its hash database of known child sexual abuse images. In addition to its frontline work, IWF provides insight on new and emerging trends and acts as a convener for key stakeholders.
- We worked with ConnectSafely—a nonprofit dedicated to educating users of connected technology about safety, privacy, and security—to develop a TikTok-specific guide for parents and teens.
- TikTok's Top 10 Tips for Families guide for the Family Online Safety Institute offers information on several tools to help teens manage how they interact with other users and who can see their videos. It includes information about privacy restrictions, content, comments, and messages.

Additionally, TikTok is part of the alliance in support of the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse. This alliance brings together governments across the world, including the United States; digital platforms and other private companies; and civil society and intergovernmental organizations. The objective is to collaborate across sectors and industries to address online threats to minors. TikTok also works with a group of non-governmental organizations as part of a CSAM intervention project designed in part to help direct users to appropriate resources.

Beyond the partnerships with leading organizations in the space, we see others finding great success in leveraging TikTok as a platform to spread critical safety messages. For example, Thorn launched an in-app campaign called NoFltr, which facilitated a conversation between youth and adults on the sharing of nude images and consent. We also worked in collaboration

with Thorn researchers and our policy team to inform our policy and features related to their findings.

We engage in these partnerships because online threats are complex and dynamic. No single company or government can solve these problems in a vacuum. We look forward to continuing our own work, as well as our collaborations, to make a meaningful difference in keeping our community safe.

TikTok's Teams and Technology

Investing in teams and technology is a core priority for me as CEO. More than 40,000 people globally work on trust and safety issues for TikTok. This includes in-house and contract moderators, as well as teams focused on safety policy, product, and operations. We also have dedicated teams focused on minor sexual exploitation, which include former law enforcement personnel who specialize in minor safety, child safety advocates and experts, and technical teams who support these efforts. TikTok invests heavily in these teams, as well as in technology to detect potential violations and suspicious accounts at scale. This year we expect to invest more than two billion dollars in trust and safety efforts, with a significant portion of that investment in our US operations.

TikTok's moderation system uses models to automatically identify videos, captions, and accounts that may contain CSAM and other violative content. Our moderation system also uses advanced technology to review comments for evidence of online grooming and other predatory behavior. And we continue to invest in developing technologies to identify new trends to help ensure we are staying ahead of the curve.

TikTok maintains a sensitive words list that we pair with models to help identify sexual/predatory texts, emojis, and phrases conveying sexual meaning or implying sexual intention. TikTok uses technology to scan for such words in video captions, optical character recognition (OCR) text, comments, and direct messages (DMs). We also employ audio event detection models, models to detect previously uploaded violative content, and keywords searches.

Each and every video uploaded to TikTok goes through automated moderation, and identified potential cases of CSAM are automatically removed or escalated for human review by a member of our moderation team. We also have models dedicated to identifying problematic accounts and behaviors. These include classifiers that identify potential grooming comments, including those requesting personal information or an offline meeting, or indicating online trading of sexually inappropriate minor content. Once such comments are identified by the classifier, they automatically will be prevented from reaching the recipient.

We aim to minimize moderators' exposure in line with industry best practices. TikTok's training materials have strict access controls and do not include visual examples of CSAM. Our specialized child safety team investigates and makes reports to NCMEC.

TikTok likewise has avenues for users to report problematic content, including CSAM. TikTok's in-app reporting function can be used to flag direct messages, comments, accounts, and videos

showing evidence of CSAM, online grooming, and other predatory behavior. These cases go to human review for further assessment and may lead to further investigations.

On top of our own technology, TikTok has integrated with NCMEC's Hash Sharing Web Services to enable the detection and removal of known CSAM at the point of upload to TikTok. The NCMEC hash list is applied against all videos uploaded to the platform and to users' avatars (profile pictures). TikTok also employs perceptual hashing techniques, such as PhotoDNA, to ensure our CSAM detection practices are robust against visual alterations (such as resizing or watermarking).

In addition, TikTok's membership in the Tech Coalition has enabled the utilization of technologies and resources such as the CSAM Keyword Hub, offered in collaboration with Thorn, to help block content or enqueue users for account review. Further, as noted above, we use IWF's keyword list to supplement our existing sensitive words list, and we deploy IWF's URL list to block URLs containing CSAM that are shared through private messages or are present in a user's bio. Finally, we are in the process of integrating with Google's Content Safety API to further support the proactive identification of never-before-seen CSAM imagery, as well as YouTube's CSAI Match to further support detection of known CSAM in videos.

When TikTok identifies CSAM or other predatory behaviors such as grooming, it is removed from the platform and reported to NCMEC. In 2021, we made 154,618 reports to NCMEC; in 2022, we made 288,125 reports. We also publish information about our content moderation decisions in our quarterly Community Guidelines Enforcement Reports. For instance, in the third quarter of 2023, 97.4 percent of our removals of content relating to youth exploitation and abuse were proactive, meaning that we took the content down before it received any reports from users or others. We are proud of the work we have done to promptly remove this content. But we also appreciate that there is no finish line when it comes to keeping our community safe, and we will never stop looking for ways to improve.

TikTok's Other Efforts to Promote a Safe Experience for Younger Users

Our commitment to minor safety goes far beyond working to protect the platform from child sexual exploitation. Safety and well-being—in particular for teens on the app—is a core priority for TikTok. To that end, we have launched several initiatives aimed at supporting teens' digital journeys and helping ensure that online experiences play a positive role in how younger users express themselves, discover ideas, and connect.

In addition to offering a range of safety and privacy controls that empower users to decide who they share content with, TikTok provides even stronger proactive protections to safeguard our teen users, and we have consistently introduced changes to support age-appropriate experiences on our platform.

For instance, accounts registered to teens under 16 are set to private by default, and their content is ineligible for recommendation to people they do not know. Additionally, teens under 16 are not able to publish their first video until they respond to a pop-up that asks them to select who may view the video. Direct messaging is not available for teens aged 13-15, and it is set by

default to "No One" for new teen users aged 16-17, requiring them to actively opt into a different sharing option.

Similarly, we limit livestream hosting to accounts registered to people 18 or older with a set number of followers. And we only allow those 18 or older to be eligible for monetization programs, such as video gifts.

TikTok imposes these limitations because we want to help teens develop positive digital habits early on. We regularly consult with leading pediatric experts, youth well-being advocates, and adolescent psychologists to develop our Youth Portal, bullying prevention guide, and other features that support youth well-being. For example, TikTok consulted with experts from the Digital Wellness Lab at Boston Children's Hospital in selecting our 60-minute default daily screen limit for minor users.

Aside from implementing age-specific restrictions, TikTok acknowledges that some users do not provide the correct information regarding their age, which is a challenge that many digital platforms face. Accordingly, in addition to our industry-standard age gating, TikTok takes a number of approaches to identify and remove accounts belonging to individuals believed to be underage. For instance, we train our trust and safety team to be alert to signs that an account may belong to a user under the age of 13. We also use other information provided by our users, such as keywords and in-app reports from our community, to help detect and remove potential underage accounts. When our safety team believes that an account may belong to an underage person, our policy is to remove the user from our platform.

Finally, TikTok recognizes the vital role that parents and guardians play in protecting the online safety of teens. With that in mind, TikTok's Family Pairing allows parents or guardians and teens to customize their safety settings based on individual needs. With Family Pairing, parents and guardians can link their TikTok accounts to their teens' accounts and set particular controls for screen time, direct messaging, search, and content, amongst others. In March 2023, we updated Family Pairing to allow for parents and guardians to receive a breakdown of their teens' screen time and to customize a schedule to mute TikTok notifications for their teens. We also started last year to prompt our community to learn more about Family Pairing, and we've reached more than 500 million people so far with this information.

TikTok's work to provide a safe and fun online environment for our teen users is never-ending. We are continuously adding to and improving these safeguards, and we welcome all feedback—both today and onward—for how we can best protect teens. We stand firm in our commitment to help ensure youth safety, and I am proud of the steps that TikTok has taken to date.

Conclusion

A fundamental component of TikTok's mission to inspire creativity and bring joy to more than 1 billion people worldwide is to continuously work toward providing a safe and secure experience on our platform. This is our commitment to our community, regardless of age. We also recognize that certain threats uniquely affect younger users, and we are resolute in working to safeguard these users' experience.

I am grateful for the opportunity to discuss these important issues with the Committee and with the leaders of other digital platforms. We are all partners in the mission of protecting our children, and I look forward to the opportunity to continue to collaborate and to work in unison on this critical effort.

HEARING BEFORE THE UNITED STATES SENATE
COMMITTEE ON THE JUDICIARY

"Big Tech and the Online Child Sexual Exploitation Crisis"

January 31, 2024

Testimony of Jason Citron,
Chief Executive Officer and Co-Founder,
Discord Inc.

Chair Durbin, Ranking Member Graham, and distinguished members of the Committee, thank you for inviting me to participate in today's hearing on this critically important topic. Ensuring child safety in the digital age demands a whole-of-society approach—parents, industry, civil society, and government all have essential roles to play.

At Discord, we believe that technology companies have an obligation to work hard to prevent our services from being exploited by bad actors who seek to harm children.

My name is Jason Citron and I am the co-founder and CEO of Discord, a voice, video, and text communications service that brings people together over shared experiences. I began my career in game programming before pursuing my own entrepreneurial ventures. Prior to co-founding Discord in 2015, I founded and served as the CEO of OpenFeint, a social mobile gaming platform. Both of my companies began as game studios with a mission to modernize the way people play games online. The success of Discord is tied inextricably to the safety of its users, and as the father of two children I understand instinctively the need to protect children's safety online. I am proud of the work that Discord is doing to prevent, detect, and respond forcefully to abuse on our service.

Allow me to start by stating clearly that Discord has a zero-tolerance policy for content or conduct that endangers or sexualizes children, and we are committed to continuing our work with parents, industry partners, and law enforcement to combat harmful content online and strengthen industry moderation practices and safety initiatives. We hope this hearing provides an opportunity for us to help the Committee understand both our overall approach to safety on Discord and the considerable work we do to prevent and address online harms targeting minors.

I. Introduction to Discord

Discord is a real-time messaging service—a chat app—for people to talk and spend time with their friends and communities online. When Discord was first launched in 2015, its initial user

base drew primarily from friends who connected through Discord's voice chat feature while playing video games. Discord has since grown to include a broad range of people who use the service to talk about a variety of topics, from sports and travel to personal finance.

As a messaging service, Discord differs from traditional social media in several important respects.

First, unlike social media platforms, Discord users are in control of their experience—they decide with whom they interact and what communities they join. There is no news feed, no endless scrolling, no counting of likes, and no "going viral." Discord is not a service designed to maximize engagement by an algorithm picking and choosing the content users see; rather, Discord emphasizes real time interaction and connection among friends.

Second, Discord's users interact with the communities they choose to join via "servers," peer-to-peer spaces that enable conversation and sharing. Servers are made up of channels where users communicate by text, voice, and video—a more organized and powerful messaging experience. Text channels allow users to interact via text-based messages, as well as images, GIFs, emojis, and other files. Voice channels allow users to communicate by voice, video, and/or livestream sharing. Users are also able to chat one-on-one via direct messages ("DMs") or in small groups via group direct messages ("GDMs"). Messages are displayed in the order in which they are sent. The vast majority of Discord servers are invite-only spaces for small groups of friends and communities. Eighty percent of communications on Discord are in smaller group servers.

Third, Discord's business model differentiates our service from traditional social media platforms. In contrast to the advertising-based business models of social media companies, Discord does not sell advertising and instead builds premium features that users can choose to purchase. For example, users can buy subscriptions to products like Nitro, which gives users the ability to upload larger files, stream video in high definition, and personalize the user experience by using custom emoji and digital stickers in conversations with others. Discord users can also purchase Server Boosts, which allow users to share these perks with members of a server to which they belong. These features are designed to improve the experience of using Discord, not to maximize engagement by keeping users on the app.

II. Discord's Users

Today, more than 150 million people log onto Discord each month. Over 75 percent of Discord's users are located outside the United States, and users communicate in more than 30 languages through the service.

Users must be at least 13 years old to create an account on Discord. We use a neutral age gate—an interface that requires new users to provide their date of birth upon creating an account. We further enforce our requirements through user reports. If we receive credible information that a given user is under 13 via a user-generated report, we ban that user from the platform until they submit a valid appeal proving that they are over 13 years old. Discord evaluates the credibility of each report and the available evidence, such as whether the user has made statements or posted information on Discord in which they admitted an age or date of birth that contradicts the information provided during account creation. If any of these factors are present, Discord will ban the user. We also welcome reports from parents or guardians who believe their child, who is under the age of 13, is active on the service.

III. Discord's Approach to Safety

Discord's mission is to bring people together around shared experiences—to create a space where everyone has a place to belong. In order to fulfill that mission, we work hard to create a safe, welcoming, and inclusive space online, especially for teens. We believe that providing online spaces where people can connect authentically with individuals who share their interests improves people's lives. To fully realize that benefit, however, service providers must establish rules and practices to prevent, detect, and remove harmful content and bad actors. Nowhere is that work more critical than in tech companies' efforts to fight the spread of child sexual exploitation.

A. Discord's Moderation Practices

Discord takes a multi-pronged approach to combating harmful content, including child sexual exploitation. This includes: (1) our robust proactive and reactive efforts to enforce our rules and remove harmful content from our services; (2) our approach to product development, whereby we implement a rigorous "safety by design" practice when developing products; (3) our policies that make clear what is and what is not allowed on the service; (4) our work with industry partners and child exploitation experts, who help Discord evaluate and improve our practices; and (5) how Discord cooperates with law enforcement.

No matter how Discord identifies violative content, users who upload child sexual abuse material (CSAM) to Discord are permanently banned from the service and reported to the National Center for Missing & Exploited Children (NCMEC), which in turn refers cases to law enforcement agencies. For incidents related to high-harm content, such as CSAM, and for violations of Discord's Terms of Service or Community Guidelines, Discord examines account identifiers to identify and ban any additional accounts the user has on the service.

1. Discord's Enforcement Work: Proactive & Reactive Efforts

Discord prioritizes enforcement against violations that present harm to our users and our service, such as child sexual exploitation, violence, and sharing of illegal content. We enforce our Community Guidelines, described in detail below, through a mix of both proactive detection and reactive measures, supported by a range of human-powered and technical solutions. Our Safety engineering teams have developed sophisticated tools to help in this work. These tools include:

Image Detection and Machine Learning: Like many of our peers, and all of the companies before the Committee today, Discord uses PhotoDNA—a tool that uses a shared industry hash database of known child sexual abuse material (CSAM)—to proactively scan images uploaded to our service in order to detect and report CSAM content and perpetrators to the National Center for Missing and Exploited Children (NCMEC), which subsequently works with local law enforcement to take appropriate action. In addition to hash matching, Discord uses internally developed tools (including machine learning techniques) and works with industry partners (including peer companies, non-profits, and researchers) to detect CSAM distribution.

In addition to our use of industry standard tools to detect and remove known CSAM, Discord invests in developing innovative technologies to proactively identify *novel* CSAM—images that have not previously been detected and so do not appear in available hash databases. We also invest in new ways to detect the exploitation of minors (commonly referred to as grooming).

- **Novel CSAM:** Recently, Discord collaborated with industry peers to build and implement a visual safety technology to detect unknown CSAM and AI-generated CSAM with positive results. Effective tools that keep children safe online should be industry standards, not a means for companies to gain a competitive advantage. In recognition of that, as of October, we have made this technology open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online.

- **Exploitation of Minors:** Text-based sexual exploitation and extortion of minors is a persistent and growing problem. We worked collaboratively with Thorn, a leading non-profit in the minor safety space, to develop a machine learning model that will detect when a minor is in communication with someone actively trying to exploit or extort them. This technology continues to be developed for broader industry use and further demonstrates our ongoing commitment not just to make Discord—but also the broader communications ecosystem—safer for kids.

In-App Reporting: Discord users are encouraged to report content they suspect to be CSAM or inappropriate sexual contact between adults and teens within DMs, GDMs, and within servers. In

order to better facilitate users' ability to report suspicious content for review, Discord streamlined the reporting process and eliminated extraneous steps between the submission of a report and review by our Trust & Safety teams. Reports concerning the highest harm material, including reports of CSAM or inappropriate contact with a minor, are prioritized for review and enforcement. In specific circumstances, reports are escalated to a dedicated investigations team, which is empowered not just to investigate specific infractions, but to conduct wider investigations and refer users to NCMEC and law enforcement agencies as appropriate.

Tools for Teens & Families: In addition to Discord's own enforcement, we provide users with sophisticated tools to help them tailor their experience and keep them safe. Fundamentally, each user has control over who they communicate with, what content they see, and what communities they join. But in recognition of the unique safety challenges that accompany teens' use of the internet, we have also created special features designed to give parents and teens even more control over their experience on Discord:

- **Family Center:** Discord provides parents with tools to better understand how their teens use our service. Our Family Center allows parents to get insights into the communities their teen children have joined, including which users their teens have recently added as friends, which users their teen has messaged or called in DMs or GDMs, and which servers their teen has joined or participated in. The Family Center also allows parents to get a weekly email with a high-level summary of the time their teen spends on Discord. The Family Center is designed not just to provide parents with greater insight into how their teen uses Discord, but to also help parents start conversations with their teen about online safety.

- **Teen Safety Assist:** In October 2023, Discord announced a series of new features designed to keep teens safe on our platform.

    - **Discord Safety Alerts** were designed in collaboration with the child safety non-profit Thorn to flag potentially unwanted conversations to teen users on Discord. This new feature helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think twice before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

    - **Sensitive Content Filters** automatically blur potentially sensitive media sent to teens in DMs, GDMs, and in servers. Blurring is enabled by default for teen users in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users.

Discord is continuing to invest in this technology and we look forward to adding even more protective messaging features for teen users in the future.

Trust & Safety Specialists: Discord's Trust & Safety experts are trained to recognize harmful content, and we have teams of specialists dedicated to reviewing categories of content that we have designated as "high-harm:" exploitative content and minor safety (which includes child sexual exploitation), extremism, and cybercrime. Our minor safety experts are trained to identify and appropriately action content and behaviors that pose risks to teens on Discord and that violate our policies prohibiting CSAM, child sexualization, and inappropriate sexual conduct with children. These teams are a key component of our child safety work and use both machine learning and human review to proactively detect bad actors and networks of bad actors before they can proliferate.

Discord's commitment to safety is reflected in our staffing decisions: more than 15 percent of our employees work on safety. By way of context, that means our Safety team is larger than our Marketing team. In addition to the significant investment we have made in our full-time staff working on these issues, we also work closely with external partners to supplement our own work on critical safety issues.

Metrics: Discord's decision to invest in proactively detecting and removing servers before they are reported to us, especially for high-harm categories, has paid dividends in user safety. In 2023 alone, Discord scanned more than 47 billion images, or roughly 130 million images every single day. During the fourth quarter of 2023, Discord proactively removed servers for child safety 96 percent of the time; servers in which CSAM was posted were removed proactively 97 percent of the time. Overall, 94 percent of servers removed for policy violations across categories during this period were removed proactively. We continue to invest in our ability to proactively detect and remove servers before they are reported to us, especially for high-harm categories.

At the account level, in the fourth quarter of 2023 Discord reported 55,955 accounts to NCMEC through our use of PhotoDNA and hashing systems such as our visual safety platform, PDQ, and CLIP. 55,638 of those reports were media (images or videos). Additionally, 317 high-harm grooming or endangerment reports were delivered to NCMEC. Discord disabled 116,219 accounts and removed 29,128 servers for child safety during this same period.

2. Product Development: Safety by Design

Because safety is critical to our core mission, Discord takes a "safety by design" approach to our work—that means a company-wide commitment to building our service safely by assessing risks, adopting mitigating measures, and evaluating metrics *throughout* the product development

cycle, not waiting until *after* a product is launched. Safety by design means much more than merely adding a "report" button that allows users to flag content for review. It also means that our teams pursue responsible product design by studying human behavior and considering the impacts of our product on user safety—before we launch it into the market.

Our process includes a risk assessment during the design phase to identify and mitigate potential safety risks. Internal experts partner with product and engineering teams to think carefully about how product features might disproportionately impact teens, and whether the product facilitates more teen-to-adult interactions and/or any other unintended harm. Our teams identify and strategize ways to mitigate safety risks with internal safety technology solutions. We also draw on our relationships with outside experts in child safety and incorporate their insights and recommendations into the product design.

Discord's safety by design approach plays a critical role in our fight against child sexual exploitation and has not only allowed the service to proactively detect, remove, and report CSAM and inappropriate sexual contact between adults and teens, but has also improved users' ability to surface concerning or suspicious content for human review. This approach, leveraging technology and human expertise, allows us to better identify and remove spaces and users on Discord that we determine to have an increased likelihood of disseminating child sexual exploitation content.

3. Discord's Policies: Community Guidelines

Discord's Community Guidelines contain the rules that everyone must follow on Discord. The Guidelines apply to all parts of the service, including content, behavior, servers, and apps. As I noted previously, Discord has a zero-tolerance policy for content or conduct that endangers or sexualizes children. This abhorrent material and behavior does not have a place on our service—or anywhere in society.

Discord's Community Guidelines prohibit CSAM and child sexualization. We do not allow CSAM on Discord, including AI-generated, photorealistic CSAM. When such imagery is found, it is removed and reported to NCMEC, which subsequently works with local law enforcement to take appropriate action. This policy is designed to ensure that the content does not proliferate and that the sexualization of children in any context is not normalized by bad actors.

Discord also has a zero-tolerance policy for inappropriate sexual conduct with children, meaning inappropriate sexual contact between adults and teens on the service, with special attention given to predatory behaviors such as online enticement and the sexual extortion of children, commonly referred to as "sextortion." This policy also applies to inappropriate sexual interactions—such as sextortion and other coercive behaviors—between a teen user and an older teen. When we

become aware of these types of incidents, we take appropriate action, including banning the accounts of offending adult users and reporting them to NCMEC, which in turn works with local law enforcement as appropriate.

Additionally, Discord has a policy on teen self-endangerment, an issue that we do not take lightly. We want teen Discord users to be able to express themselves freely while also taking steps to ensure that they don't engage in risky behaviors that could endanger their safety and well-being. In order to help teen users stay safe, our policies state that those under the age of 18 are not allowed to send or access any sexually explicit content, and users are only allowed to post explicit content in spaces that are behind an age-gate or in DMs or GDMs with users who are 18 and older. We also believe that teens dating online can result in self-endangerment. Under this policy, teen dating servers are prohibited on Discord, and we will take action against users who are engaging in this behavior.

Discord's Community Guidelines also prohibit a much broader range of harmful content and behavior, including content that glorifies, promotes, or seeks to normalize suicide and other acts of physical self-harm; illegal behavior; threats and harassment; hate speech and hateful conduct; violent extremism, including the glorification of violence; and other categories of harmful content and behavior that have no place on our service. Users seeking to better understand our approach to safety can learn more about the Community Guidelines in the Discord Policy Hub, a resource for users that houses in-depth policy explainers across a range of topics, including minor safety.

Our work in this space is constantly evolving to meet the changing threat landscape, and our Platform Policy team continues to monitor new developments in order to update and adjust our policies as needed.

4. External Engagement & Partnerships

Discord's work to halt the spread of harmful content does not take place in a vacuum; we work closely with industry partners and outside experts to inform and enhance our efforts in this space. We develop and share best practices and information with peer companies, non-governmental organizations, academics, and subject matter experts.

To enhance our efforts at fighting child sexual exploitation specifically, Discord works with a range of groups to assist in the detection of harmful content and on the improvement of our internal policies and processes, including:

- NCMEC,
- INHOPE,
- Thorn,

- Family Online Safety Institute,
- Internet Watch Foundation, and
- Technology Coalition (a group of companies working together to end online child sexual exploitation and abuse), among others.

We also work closely with nearly 50 organizations around the world as part of our Safety Reporting Network, with special consideration for child safety. Trusted reporters are outside experts—civil society groups, non-governmental organizations—that help alert us to indicators of harmful content and assist in our efforts to detect and action such content on Discord.

In 2023, Discord was a founding member of Lantern, a cross-platform signal sharing program organized by the Tech Coalition that strengthens how companies enforce their child safety policies. Lantern allows tech companies to securely and responsibly share signals about activities and accounts that violate our child safety policies. As a result of Discord's participation in the program, we have been able to scale our enforcement against offending accounts and share data related to the highest-harm abuses with other participating companies. Discord has also acted on data shared with us through this program. The cross-company cooperation facilitated by Lantern helps prevent predatory actors from evading detection across services, and Discord is proud to participate in this program.

5. Cooperation with Law Enforcement

Discord provides law enforcement agencies with necessary information while respecting the privacy and rights of our users. When we receive a request from a law enforcement agency seeking information about a Discord user, we expeditiously review it to ensure legal compliance and its overall validity. Discord may also disclose user data to law enforcement in emergency situations when we possess a good faith belief that there is an imminent risk of serious physical injury or death. In order to ensure requests are handled appropriately and expeditiously, we adopted clear guidelines for law enforcement authorities seeking information or records about Discord users.

Discord publishes quarterly Transparency Reports that detail our interactions with law enforcement and provide greater insight into the work we do to keep users safe. Transparency is one of Discord's core safety values and we view the publication of these reports as a way to honor our commitment to that value.

We evaluate each law enforcement request that Discord receives to ensure its validity and provide responsive information. According to our most recent Transparency Report issued on January 16, 2024, Discord received 1,326 pieces of legal process from federal and state authorities in the United States during the fourth quarter of 2023, finding 1,073 (81 percent) both

legally valid and sufficiently specific for us to identify an account and produce relevant information. Discord received 1,125 preservation requests globally during the fourth quarter of 2023, finding 813 (72 percent) valid and specific enough for us to identify an account and preserve relevant information.

Discord received 203 valid emergency disclosure requests from law enforcement during this period. These requests originated from law enforcement agencies around the world. Discord reviews each emergency request to authenticate the requestor, confirm the request is specific enough for us to identify an account, and verify that the request meets the legal requirements necessary for providers to disclose information in emergency situations. Discord disclosed information in response to 159 (78 percent) emergency disclosure requests.

IV. Conclusion

Discord's mission is to bring people together around shared experiences. We view user safety not only as our responsibility but absolutely essential to fulfilling our mission and creating a space online where everyone has a place to belong.

Discord is committed to identifying new and innovative ways to halt the spread of online child sexual exploitation. Although we are a mid-sized service, we have invested heavily in staffing, technological innovation, and partnerships to help ensure the safety of our users and the health of our service. As a result, since Discord was founded in 2015, we have banned more than 3 million accounts for child safety violations and reported more than 150,000 accounts to NCMEC. We have also worked closely with our partners in law enforcement as appropriate and made it a priority to engage efficiently and productively when agencies flag potential safety concerns.

The proliferation of child sexual exploitation content on the internet requires that industry work together with parents, civil society, and government to create new technologies and explore innovative methods to prevent, detect, and halt online harms to children. Discord is committed to continuing our work to combat harmful content online and strengthen industry moderation practices and safety initiatives. I am eager to learn more about the Committee's concerns and to find ways that Discord can be a partner in identifying solutions.

Testimony of Evan Spiegel
Co-Founder and CEO, Snap Inc.

Hearing before the United States Senate Committee on the Judiciary
January 31, 2024

Chairman Durbin, Ranking Member Graham, and members of the Committee, thank you for inviting me to appear today to update you on our efforts to protect young people on Snapchat. I'm Evan Spiegel, the co-founder and CEO of Snap. Our service, Snapchat, is used by over 100 million Americans, including more than 20 million teenagers, to communicate with their friends and family. We have an enormous responsibility to keep our community safe.

We know that the scale and widespread use of Snapchat mean that bad actors will try to abuse our service and take advantage of our community. That's why we are constantly improving our safety tools and investing in protecting our community from the ever-evolving threat landscape. Protecting Snapchatters is our moral responsibility and a business imperative. I want to share more about some of the biggest threats we are working to combat, but first I want to provide a bit of background about our service as this is my first time appearing before the Committee.

When my co-founder Bobby Murphy and I first built Snapchat in 2011, we wanted something different. We grew up with social media and it made us feel miserable – a public, permanent, popularity contest filled with constant judgment. Social media was for perfect pictures, instead of the everyday moments that we believe strengthen real friendships.

We built Snapchat to offer a new way to communicate with friends and family, to share the moment, in the moment, and help people feel together even if they are physically far apart. On average, people spend most of their time on Snapchat talking with their friends. We designed Snapchat to open into the camera, instead of a content feed, to encourage creativity instead of passive consumption. When people share their Story with friends on Snapchat there are no public likes or comments.

By embracing ephemerality, and deleting messages by default, we gave Snapchat the lightness of a phone call or face-to-face conversation that isn't recorded or saved forever. This has helped millions of Americans feel more comfortable expressing themselves and sharing how they really feel with their friends and family. When people sign up for Snapchat, we make it clear that even though conversations are deleted by default, messages can be easily saved or screenshot by the recipient.

When we build new features, we make business trade-offs to better serve our community and help keep Snapchat safe. For example, when we built our content service, we decided to proactively moderate content before it can be broadly distributed to help prevent the spread of harmful content. We also pay media publishers and creators a share of our revenue to

incentivize them to produce content that is entertaining and consistent with our content guidelines.

We've designed our service to require communication between friends to be opt-in, meaning people have to proactively choose who they communicate with, unlike text messaging, where any stranger can message someone if they have their phone number. Friend lists are private on Snapchat, which not only reduces social pressure but also limits the ability of predators to find a person's friends on Snapchat.

We want Snapchat to be safe for everyone, and we offer extra protections for minors to help prevent unwanted contact and provide an age-appropriate experience. Snapchat's default "Contact Me" settings are set to friends and phone contacts only for all accounts, and can't be expanded. If a minor receives a friend request from someone they don't share a mutual friend with, we provide a warning before they start communicating to make sure it is someone they know. As a result, approximately 90% of friend requests received by minors on Snapchat are from someone with at least one mutual friend in common. Our goal is to make it as difficult as possible for people to be contacted by someone they don't already know.

We encourage Snapchatters to report unwanted contact or violating content, and we block the offending account. For people who do not have a Snapchat account but want to make a report, we also offer reporting tools on our website. All reports are confidential and our Trust and Safety team works 24 hours per day, seven days per week, around the world to review each report and consistently enforce our rules.

When we take action on illegal or potentially harmful content, we retain the evidence for an extended period, which allows us to support law enforcement in their investigations. We also proactively escalate to law enforcement any content that appears to involve imminent danger of death or serious physical injury and typically respond to emergency data disclosure requests within 30 minutes. We want criminals who abuse Snapchat to be brought to justice.

There are three major threats to our community that we are working to eliminate from our service: extortion, the distribution of child sexual abuse material, and illicit drugs.

The first is the concerning rise in financially-motivated sextortion, a form of blackmail where criminals pose as a potential love interest and convince victims to send compromising images. The bad actors then threaten to release the images and demand payment, often in the form of gift cards, which can be photographed and shared via chat. Many of these cases involve predators located outside of the United States which makes enforcement through the legal process more challenging.

In response to this growing crisis, we have developed new tools to proactively detect these bad actors on our service and seek to intervene before the conversation can escalate to extortion. When harassment or sexual content is reported to us by our community, our team acts quickly, usually taking action within 15 minutes.

Second, we are also identifying criminals who seek to re-victimize children who have been sexually abused by sharing images and videos of the abuse on our service. We scan image and video uploads to Snapchat for known child sexual abuse material and report it to the National Center for Missing and Exploited Children. In 2023 we made 690,000 reports that led to more than 1,000 arrests. We do not anticipate implementing encryption in a way that would prevent us from scanning uploads for known child sexual abuse imagery.

Third, is the ongoing and devastating fentanyl epidemic that claimed the lives of over 100,000 Americans last year. We are determined to remove drug dealers and drug-related content from our service. We proactively scan our service for illegal drug content, disable drug dealer accounts and ban their devices from accessing our service, preserve the evidence, and make referrals to law enforcement, including the Drug Enforcement Administration. In 2023, we removed more than 2.2 million pieces of drug-related content, disabled the 705,000 related accounts, and blocked the devices associated with those accounts from using Snapchat.

We block drug-related search terms and redirect people searching for drugs to educational materials on our service. Fentanyl poses a unique threat, because it is incredibly lethal and laces nearly every type of drug and counterfeit pill available on the street. That's why we believe education is so important and we have invested in public awareness campaigns, such as One Pill Can Kill, which was viewed over 260 million times on Snapchat, and the Ad Council's Real Deal on Fentanyl to educate our community on the dangers of counterfeit pills.

In addition to the parental controls available as part of the iOS and Android operating systems, we have worked to empower parents by giving them more tools to supervise the way their teens use Snapchat. Parents can use our Family Center to view a list of people that their teen is communicating with using our service. This resembles how we believe parents monitor their teens' activity in the real world – where parents want to know who their teens are spending time with but don't need to listen in on every private conversation. Family Center also allows parents to review privacy settings and set content controls.

I hope this hearing represents an opportunity to move forward important legislation like the Kids Online Safety Act and the Cooper Davis Act. We support this legislation, not only in word, but in deed, and we have worked to ensure our service lives up to the legislative requirements before they are formal, legal obligations. This includes limiting who can communicate with teens to friends and contacts only, offering in-app parental tools, proactively identifying and removing harmful content, and referring lethal drug content to law enforcement. We are continuing to work with the Committee on the Stop CSAM Act, which we believe represents meaningful progress towards eradicating child sexual exploitation from online services.

Many of the largest and most successful Internet companies today were born here in the United States of America, and we must lead not only in technical innovation but also in smart regulation. That is why we support a comprehensive federal privacy bill that will protect the data privacy of all Americans and create consistent privacy standards for all online services.

I want to take this opportunity to express our heartfelt appreciation for all of the incredible partners and collaborators we work with across the industry, in government, and the nonprofits

and NGOs who share our goal of keeping our community, and especially young people, safe. We are particularly grateful to law enforcement and the first responders who are vital to these efforts. For the sake of brevity and for fear of leaving someone out, I won't list everyone individually, but please accept our deepest thanks and utmost gratitude.

We consistently hear from our community that using Snapchat makes them feel happy and we know that relationships with friends and family are important for mental health and wellbeing. We recently commissioned research from the National Opinion Research Center at the University of Chicago that found respondents who use Snapchat report higher satisfaction with the quality of their friendships and relationships with family than non-Snapchatters. Our deep desire to make a positive impact in the world motivates us every day to make sure that our service is used in a safe and healthy way.

Fundamentally, we believe that online interaction should be safer than offline interaction. While we recognize that it may be virtually impossible to eliminate all of the risks involved with using online services, we are determined to do our part to protect the Snapchat community. Young people represent our country's future and we must work together to protect them.

Thank you.

###

Written Testimony of Linda Yaccarino

CEO, X Corp.

United States Senate Committee on the Judiciary

Hearing on Big Tech and the Online Child Sexual Exploitation Crisis

January 31, 2024

Chairman Durbin, Ranking Member Graham, and esteemed members of the Committee, thank you for the opportunity to discuss X's work to protect the safety of minors online.

X has zero tolerance for Child Sexual Exploitation (CSE), and we are determined to make X inhospitable for actors who seek to exploit minors. In 2023, we made clear that our top priority was tackling CSE online.

As an entirely new company, X has strengthened its policies and enforcement to tackle CSE. We are taking aggressive action against users that distribute CSE and the networks of users who engage with this horrible content.

While X is not the platform of choice for children and minors – users between 13-17 account for less than 1% of our U.S daily users – we have made it more difficult for bad actors to share or engage with CSE material on X, while simultaneously making it easier for our users to report CSE content.

We are improving our detection mechanisms to find more reportable content on the platform to report to the National Center for Missing and Exploited Children (NCMEC). In addition, we are building a Trust and Safety center of excellence in Austin, Texas, to hire more in-house agents so we can continue to accelerate our impact.

The following is a comprehensive update on our progress and our continued investment in this critical area.

CSE Actions

In 2023, as a result of our investment in additional tools and technology to combat CSE, X suspended 12.4 million accounts for violating our CSE policies. This is up from 2.3 million accounts in 2022.

Along with taking action under our rules, we also work closely with NCMEC. In 2023, X sent 850,000 reports to NCMEC, including our first ever fully-automated report. This is over eight times more than Twitter sent in 2022.

Not only are we detecting more bad actors faster, we are also building new defenses that proactively reduce the discoverability of posts that contain this type of content.

Advanced technology and proactive monitoring

We are investing in products and people to bolster our ability to detect and action more content and accounts, and are actively evaluating advanced technologies from third-party developers that can enhance our capabilities. Some highlights include:

- Automated NCMEC reporting: In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and

report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year.
- Expanded Hash Matching to Videos and GIFs: For the first time ever, we are evaluating all videos and GIFs posted on X for CSAM. Since launching this new approach in July 2023, we have matched over 70,000 pieces of media.
- Launched Search Intervention for CSE Keywords: CSAM impressions occur more on search than on any other product surface. In December 2022, we launched the ability to entirely block search results for certain terms. We have since added more than 2,500 CSE keywords and phrases to this list to prevent users from searching for common CSE terms.

Stronger partnerships, global cooperation

We are closely working with trusted organizations that are aligned with our mission to combat online CSE.

Foundational to our work is our multidimensional partnership with NCMEC, which manages the CyberTipline program, regularly convenes global stakeholders and facilitates actionable feedback from law enforcement that makes us better. Other instrumental partners are the Tech Coalition and WeProtect, alliances that push our innovation and provide critical information sharing on emerging threats and behaviors.

In December 2022, we launched a new product partnership that allows us to take down more CSAM than before. Built by Thorn, Safer allows tech platforms to identify, remove, and report child sexual abuse material at scale.

The Internet Watch Foundation, where we have been a member since 2014, provides technical signals, trend and threat analysis and member collaboration opportunities that directly impact our ability to remediate CSAM content. International hotline operators like Point de Contact flag alleged illicit content on the basis of reports and their own monitoring. We are an active participant in the Child Protection Lab of the Paris Peace Forum answering the international call to stand up for children's rights in the digital environment.

Ongoing improvements and regular review

At X, we are more vigilant and aggressive than ever in our enforcement. Our team regularly reviews and implements improvements to the measures we take to combat online child sexual exploitation to ensure their ongoing efficacy and performance. Our increased investment in this area throughout the year has yielded significant, measurable results.

Since April, we have increased training for content moderators on the tools and policies for NCMEC reporting. In turn, this has led to a 10x increase in the volume of manually-submitted NCMEC reports, from an average of 6,300 reports per month to an average of 64,000 reports per month from June through November 2023.



Working with law enforcement

Ultimately, it is critical that the bad actors be brought to justice and that law enforcement has the tools and resources they need to prosecute these heinous crimes. X cooperates with law enforcement around the world and provides an online portal to submit removal requests, information requests, preservation requests, and emergency requests.

Looking forward

Our work to stop child sexual exploitation will never stop.

In 2024, we will continue our strong investment in this critical area and expand our efforts to educate our users about the importance of helping us to combat child sexual exploitation online. We are committed to making X a place where freedom of expression and users' safety are not compromised in the public conversation.

Child Sexual Exploitation (CSE) Policies

We have a zero-tolerance child sexual exploitation policy on X.
X has **zero tolerance towards any material that features or promotes child sexual exploitation**, one of the most serious violations of Our Rules. This may include media, text, illustrated, or computer-generated images. Regardless of the intent, viewing, sharing, or linking to child sexual exploitation material contributes to the re-victimization of the depicted children. This also applies to content that may further contribute to victimization of children through the promotion or glorification of child sexual exploitation. For the purposes of this policy, a minor is any person under the age of 18.

What is in violation of this policy?

Any content that depicts or promotes child sexual exploitation including, but not limited to:

- visual depictions of a child engaging in sexually explicit or sexually suggestive acts;
- illustrated, computer-generated or other forms of realistic depictions of a human child in a sexually explicit context, or engaging in sexually explicit acts;
- sexualized commentaries about or directed at a known or unknown minor; and
- links to third-party sites that host child sexual exploitation material.

The following behaviors are also not permitted:

- sharing fantasies about or promoting engagement in child sexual exploitation;
- expressing a desire to obtain materials that feature child sexual exploitation;
- recruiting, advertising or expressing an interest in a commercial sex act involving a child, or in harboring and/or transporting a child for sexual purposes;
- sending sexually explicit media to a child;



- engaging or trying to engage a child in a sexually explicit conversation;
- trying to obtain sexually explicit media from a child or trying to engage a child in sexual activity through blackmail or other incentives;
- identifying alleged victims of childhood sexual exploitation by name or image; and
- promoting or normalizing sexual attraction to minors as a form of identity or sexual orientation.

Who can report violations of this policy?

Anyone can report potential violations of this policy, whether they have an X account or not.

In addition to reporting Child Sexual Abuse Material (CSAM), you have the option to report any issues related to Child Safety including Child Sexual Exploitation, grooming (of a minor), and Physical Child Abuse in the X app. Find detailed instructions on how to report on our Help Center.

What happens if you violate this policy?

In the majority of cases, the consequence for violating our child sexual exploitation policy is **immediate and permanent suspension**. In addition, violators will be prohibited from creating any new accounts in the future. Note: when we're made aware of content depicting or promoting child sexual exploitation, including links to third party sites where this content can be accessed, they will be removed without further notice and reported to the **National Center for Missing & Exploited Children (NCMEC)**.

In a limited number of situations, where we haven't identified any malicious intent, we will require you to remove this content. We will also temporarily lock you out of your account before you can Post again. Further violations will lead to your account being permanently suspended. If you believe that your account was suspended in error, you can submit an appeal.

Non-Consensual Nudity Policy

You may not post or share intimate photos or videos of someone that were produced or distributed without their consent.

Sharing explicit sexual images or videos of someone online without their consent is a severe violation of their privacy and the X Rules. Sometimes referred to as revenge porn, this content poses serious safety and security risks for people affected and can lead to physical, emotional, and financial hardship.

What is a violation of this policy?

Under this policy, you can't post or share explicit images or videos that were taken, appear to have been taken or that were shared without the consent of the people involved.

Examples of the types of content that violate this policy include, but are not limited to:

- hidden camera content featuring nudity, partial nudity, and/or sexual acts;
- creepshots or upskirts - images or videos taken of people's buttocks, up an individual's skirt/dress or other clothes that allows people to see the person's genitals, buttocks, or breasts;
- images or videos that superimpose or otherwise digitally manipulate an individual's face onto another person's nude body;
- images or videos that are taken in an intimate setting and not intended for public distribution; and
- offering a bounty or financial reward in exchange for intimate images or videos.

What happens if you violate this policy?

We will **immediately and permanently suspend any account that we identify as the original poster of intimate media** that was created or shared without consent. We will do the same with any account that posts only this type of content, e.g., accounts dedicated to sharing upskirt images.

In other cases, we may not suspend an account immediately. This is because some people share this content inadvertently, to express shock, disbelief or to denounce this practice. In these cases, we will require you to remove this content. We will also temporarily lock you out of your account before you can post again. If you violate this policy again after your first warning, your account will be permanently suspended. If you believe that your account was suspended in error, you can submit an appeal.

HEARING BEFORE THE UNITED STATES SENATE

COMMITTEE ON THE JUDICIARY

January 31, 2024

Testimony of Mark Zuckerberg
Founder and Chief Executive Officer, Meta

I. Introduction

Chairman Durbin, Ranking Member Graham, and members of the Committee:

Every day, teenagers and young people go online to stay connected to their friends and family, find community, and get support. Teens do amazing things on our services. They use our apps to feel more connected, informed, and entertained, as well as to express themselves, create things, and explore their interests. Overall, teens tell us this is a positive part of their lives. But some still face challenges online, and we work hard to provide support and controls to reduce potential harms.

Being a parent is one of the hardest jobs in the world. Technology gives us new ways to communicate with our kids and feel connected to their lives, but it can make parenting more complicated, too. It's important to me that our services are positive for everyone who uses them. We're focused on building controls to help parents navigate the reality of raising kids today, including tools that enable them to be more involved in their kids' decisions.

We want teens to have safe, age-appropriate experiences on our apps, and we want to help parents manage those experiences. That's why in the last 8 years we've introduced more than 30 different tools, resources, and features to help parents and teens. These include controls that let parents set limits on when and for how long their teen can use our services, see who they're following, and know if they've reported anyone who might be bullying them. For teens, these tools include nudges that remind them when they've been using Instagram for a while or when it's late and they might want to go to sleep, and the ability to hide words, topics, or people from their experience without those people finding out.

With so much of our kids' lives spent on mobile devices and social media, it's important to ask and think about the effects on teens—especially on mental health and well-being. This is a critical issue, and we take it seriously. Mental health is a complex issue, and the existing body of scientific work has not shown a causal link between using social media and young people having worse mental health outcomes. A recent report from the National Academies of Sciences evaluated results from more than 300 studies and determined that the research "did not support the conclusion that social media causes changes in adolescent mental health at the population level." It also suggested that social media can provide significant positive benefits when young people use it to express themselves, explore, and connect with others. We'll continue to monitor research in this area and remain vigilant against any emerging risks.

Keeping young people safe online has been a challenge since the start of the internet. As threats from criminals evolve, we have to evolve our defenses. We work closely with law enforcement to find and stop bad actors. Still, no matter how much we invest or how effective our tools are, this is an adversarial space. There is always more to learn and more improvements to make. We remain ready to work with members of this Committee, the industry, and parents to strengthen our services and make the internet safer for everyone.

I'm proud of the work our teams have done to improve online child safety, not just on our services but across the entire internet. We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone. We've built and shared tools for removing bad content across the internet, and we look at a wide range of signals to detect problematic behavior. We go beyond legal requirements and use sophisticated technology to proactively seek out abusive material, and as a result, we find and report more inappropriate content than anyone else in the industry. As the National Center for Missing and Exploited Children (NCMEC) put it just this week, Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs."

I hope we can have a substantive discussion that drives improvements across the industry, including new legislation that delivers what parents say they want most: a clear system for age verification and parental control over what apps their kids are using. For example, 3 out of 4 parents favor introducing app store age verification, and 4 out of 5 parents want legislation requiring app stores to get parental approval whenever teens download apps. We support this. Parents of teens under 16 should have the final say on what apps are appropriate for their children, and this approach would leverage the parental approval system for purchases that app stores already provide today, so there'd be no need for parents and teens to share a government ID or other personal information with every one of the thousands of apps out there. We're also in favor of setting industry standards on age-appropriate content and limiting signals for advertising to teens to age and location, not behavior. We're ready to work with any member of this Committee who wants to discuss legislation in these areas and any of our peers across the industry to help move this forward.

II. Our Work

Teen well-being and child safety are extremely important to us. We have many teams dedicated to these issues, and we lead the industry in a lot of the areas we're here to discuss.

We've built more than 30 tools, resources, and features to help protect teens and give parents oversight and control over how teens are using our services, including:

- Parental supervision tools, which let teens or their parents set daily limits for the total time that teens can spend on Instagram, Facebook, Messenger, Quest, and Horizon. Teens and parents can also set scheduled breaks that block access during specific hours of the day, such as during school or dinner time. So far, over 90% of U.S. teens are still using daily limits 30 days after initial adoption.

- Take A Break notifications, which show full-screen reminders to leave the Instagram app.

- Prompting teens to turn on Quiet Mode, which turns off notifications and auto-replies to messages if they're on the app for a specific amount of time at night.

- Nudges, which include alerts that notify teens that it might be time to look at something different if they've been scrolling on the same topic for a while, or that it's getting late and might be time to close the app for the night.

- Age verification technology on Instagram to confirm a teen's age when they change their birthday from under 18 to over 18.

We also provide special protection for teen accounts:

- Accounts for people under 16 (or under 18 in certain countries) are defaulted to private, so teens can control who sees or responds to their content.

- Teens are defaulted into the most restrictive content and recommendations settings to make it more difficult to come across potentially sensitive content or accounts. 99% of teens who are defaulted globally and in the U.S. are still using this setting a year later.

- We recently announced additional steps to help protect teens from unwanted contact, turning off their ability to receive DMs from anyone they don't follow or aren't connected to on Instagram—including other teens—by default.

- We prompt teens to review and restrict their privacy settings.

- We offer the option to hide like counts, so people don't have to show others like counts on their own posts or see likes on other people's posts.

- In addition to these teen-specific protections, we hide results for searches for terms related to suicide, self-harm, and eating disorders, instead offering access to expert resources for everyone on Instagram.

Parents and guardians know what's best for their teens, so we also make it easy for them to be involved in their teens' online experiences with supervision tools and expert-backed resources:

- Parents can decide when, and for how long, their teens use Instagram, see who their teens are following, and receive reports when they block someone or report something.

- On Facebook, parents can see insights like time spent, schedule breaks for their teens, and access expert resources on managing their teens' time online.

- Over 90% of guardians and teens in the U.S. who choose supervision experiences on Facebook or Instagram are still using them 30 days after initial adoption.

- We've implemented similar parental supervision tools across our apps.

We've built tools and policies specifically to help young people manage interactions with adults:

- As noted above, we turn off teens' ability to receive messages from anyone they don't follow or aren't connected to on Instagram by default. If a teen is already connected with a potentially suspicious adult, we send the teen a safety notice.

- We restrict adults over the age of 19 from messaging teens who don't follow them, and we limit the type and number of direct messages people can send to someone who doesn't follow them to one text-only message.

- We use prompts or safety notices to encourage teens to be cautious in conversations with adults they're already connected to, and give them an option to end the conversation, or to block, report, or restrict the adult.

- We've made it easier to report content with a new dedicated option to prioritize a report if it "involves a child" on Facebook and Instagram.

We build technology specifically to help tackle some of the most serious online risks, and we share it to help our whole industry get better:

- We built technology behind Project Lantern, the only program that allows apps to share data about people who break child safety rules.

- We were a founding member of Take It Down, the service that enables young people to prevent their nude images from being spread online. This is an important tool that a teen can use to protect against the threat of sextortion.

- In 2020, we joined Google, Microsoft, and 15 other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual abuse.

- We work closely with safety advisors and professionals, as well as leading online safety nonprofits and NGOs to combat child sexual exploitation and aid its victims.

- We've partnered with child-safety organizations and academic researchers to complete child-safety research that has helped move the industry forward. For example, we recently partnered with the Center for Open Science on a pilot program to share privacy-preserving social media data with academic researchers to study well-being.

We also work to find, remove, and report child sexual abuse material and disrupt the networks of criminals behind it:

- We developed technology that identifies potentially suspicious adults, reviewing over 60 signals to proactively find and restrict potential predators. We deploy machine learning to proactively detect accounts engaged in certain suspicious patterns of behavior by analyzing dozens of combinations of metadata and public signals, such as if a teen blocks or reports an adult.

- When we identify these accounts, we limit their ability to find, follow, or interact with teens or each other, and we automatically remove them if they exhibit a number of these signals.

- As required by law, we report all apparent instances of child exploitation identified on our site from anywhere in the world to NCMEC, which coordinates with law enforcement authorities from around the world.

- We respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations. We provide operational guidelines to law enforcement who seek records from Facebook or Instagram.

- Between 2020 and 2023, our teams disrupted 37 abusive networks and removed nearly 200,000 accounts associated with those networks.

- In Q3 2023, we removed 16.9 million pieces of child sexual exploitation content on Facebook and 1.6 million pieces on Instagram.

- In Q3 2023, of the child sexual exploitation content we actioned, we detected 99% on Facebook and 96% on Instagram before it was reported by our users.

III. Our Commitment

We want everyone who uses our services to have safe, positive, and age-appropriate experiences, and we approach all our work on child safety and teen mental health with this in mind. We build comprehensive controls into our services, we work with parents, experts, and teens to get their input, and we engage with Congress about what else needs to be done.

We're committed to protecting young people from abuse on our services, but this is an ongoing challenge. As we improve defenses in one area, criminals shift their tactics, and we have to come up with new responses. We'll continue working with parents, experts, industry peers, and Congress to try to improve child safety, not just on our services, but across the internet as a whole.

That goes for our work on youth well-being and mental health, too. We'll continue to study this ourselves, monitor external studies, and open up our data for academic researchers, and we'll keep working on additional tools and resources that give parents and teens more control over their experiences online. I look forward to discussing these important issues with you today.

TikTok's Responses to Questions for the Record from Senator Durbin Submitted March 7, 2024

1. **For each year from 2019 to 2023, please provide the following:**

    a. **the total number of users on your platform;**

TikTok is an online community of more than a billion people worldwide -- including well over 170 million Americans. We believe that the average U.S. user is over 30 years of age.

    b. **the total number of users under the age of 18 on your platform;**

As a privately held company, TikTok does not disclose specific user numbers.

    c. **the estimated number of users under the age of 13 on your platform;**

As a privately held company, TikTok does not disclose specific user numbers.

    d. **the number of users of your platform under the age of 18 who were paired with a parent or guardian's account using your Family Pairing tool?**

As a privately held company, TikTok does not disclose detailed statistics about its users or adoption of specific features.

    e. **your company's annual revenue;**

As a privately held company, TikTok does not disclose annual revenue or other financial specifics.

    f. **your company's annual budget for trust and safety;**

We expect to invest more than two billion dollars in trust and safety efforts in 2024, with a significant part of that investment in our U.S. operations.

    g. **your company's annual budget to address online child sexual exploitation;**

As a privately held company, TikTok does not disclose annual revenue or other financial specifics.

    h. **the total number of employees working to address trust and safety;**

TikTok currently has more than 40,000 trust and safety professionals working to protect our community.

i. the total number of employees working to address online child sexual exploitation.

TikTok currently has more than 40,000 trust and safety professionals working to protect our community. We deploy a combination of automated technology and skilled human moderators who take into account additional context and nuance which may not always be picked up by technology. Our moderators are trained across all Community Guidelines violations, including youth safety issues. Some violations require further work by specialized moderators in specialized queues.

2. How did your company determine that 13 was the appropriate age for a child to begin using your platform?

TikTok is deeply committed to ensuring that its platform provides a safe and positive experience, especially for people under the age of 18. TikTok follows the Federal Trade Commission's ("FTC") guidance and provides a separate experience in the U.S. for users under 13. This curated viewing experience has additional safeguards and privacy protections designed specifically for a younger audience. In line with the FTC's guidance on children's privacy, users 13 and older may access the full TikTok experience.

TikTok provides additional safeguards for teen users, including: (1) a default daily screen time limit of 60 minutes, (2) setting accounts to private by default, (3) restricting access to host LIVE content or engage in financial transactions, and (4) turning off push notifications at night. Users under 16 may not have content recommended to people they do not know in the For You feed, may not send or receive direct messages, may only receive comments on their content from friends, and their videos are not available for duets or stitches. These measures help further TikTok's goal of providing young people with an experience that is developmentally appropriate and help ensure a safe space for self-exploration.

3. What legal obligation does your company have in the United States to ensure that your platforms are safe for children before they are launched?

TikTok's goal of providing an age-appropriate experience to its users begins with an industry-standard neutral age gate that is consistent with the Federal Trade Commission's ("FTC") guidance for age verification under the Children's Online Privacy Protection Act ("COPPA"). If an individual selects a birthdate that indicates that they are under the age of 13 when creating a TikTok account in the U.S., they are directed to TikTok's under 13 experience, where they can watch a curated library of age-appropriate videos. In addition to being restricted to only certain approved content, users in the under 13 experience cannot access many of the features and functions that are available to users on the 13+ experience. For example, they are not able to post videos on the platform, comment on videos, message other users, maintain a profile or followers, receive ads, or be directed off the TikTok platform.

Beyond age gate, TikTok uses technologies and human moderators, as well as user and third party reporting, to detect and remove users in the 13+ experience who are suspected to be under 13. An account in the 13+ experience that is flagged as being potentially under 13 is routed to a

dedicated team of trained moderators who would review the account to determine if it should be banned for not meeting the minimum age requirement. If the moderator makes a determination that the account belongs to a suspected underage user, the account would be removed from the 13+ experience.

4. **For users under the age of 18,**
    a. **what are the default privacy settings for their accounts?**
    b. **what limitations are placed by default on content these users can access, content that will be directed toward them, and individuals they can communicate with?**
    c. **can they change their default settings without the awareness of their parent or guardian, or without the consent of their parent or guardian?**
    d. **in 2023, how many changed their default settings?**

TikTok promotes a safe and age-appropriate experience for teen users between ages 13 through 17 by utilizing a multi-faceted, multi-level approach that currently includes the following age-based default settings:

- Accounts ages 13-15:
    - Have a 60 minute screentime limit by default
    - Cannot send or receive virtual gifts[1]
    - Are set to private by default
    - Cannot access direct messaging
    - Cannot host a livestream
    - Cannot buy or sell on TikTok Shop
    - Are not able to have their content recommended to people they do not know in the For You feed
    - Cannot have their content Dueted or Stitched
    - Do not receive push notifications from 9 pm to 8 am
    - Do not have their accounts recommended to others by default. If the user changes this option, their account still will not be recommended to people 18 and over
- Accounts ages 16-17:
    - Have a 60 minute screentime limit by default
    - Cannot send or receive virtual gifts
    - Private account option is pre-selected by default at account registration

[1] https://newsroom.tiktok.com/en-us/updating-our-gifting-policies

- Cannot host a livestream[2]
- Cannot buy or sell on TikTok Shop
- Do not receive push notifications from 10 pm to 8 am
- Do not have their accounts recommended to others by default. If the user changes this option, their account still will not be recommended to people 18 and over

In addition, we offer parental controls. Family Pairing allows a parent or guardian to link their TikTok account to their teen's account to directly manage a number of safety controls for their teen's account, including:

- Account privacy: Set their teen's account to private or public.
- Comments: Restrict who can comment on their teen's videos.
- Direct Messages: Restrict who, if anyone, can send private messages to the account
- Daily screen limits: Set the amount of time spent on TikTok each day
- Screentime dashboard: See their teen's time spent on TikTok
- Search settings: Restrict their ability to search for content
- Push notifications: Restrict push notifications to their teens during certain times
- Keyword filters: Add hashtags or keywords they would prefer their teen not see in their For You feed recommendations
- STEM feed: Enable TikTok's STEM feed - a feed featuring videos related to science, technology, engineering, and math - on their teen's account

We help our community understand and control how they spend their time on TikTok. For example, we offer:

- Screen time dashboards that provide insight into how and when a community member is using TikTok;
- Screen time breaks that nudge our community members to take a break from the app after a period of uninterrupted screen time;
- Sleep reminders that allow people to set a reminder to log off at a certain time of day;
- Daily screen time limits that allow people to determine on how much time they spend on TikTok each day; and
- Screen time updates that allow people to receive weekly info about their screen time usage.

[2] https://newsroom.tiktok.com/en-us/enhancing-the-live-community-experience-with-new-features-updates-and-policies

TikTok has also introduced Content Levels[3] to help prevent content with overtly mature themes, such as graphic violence or cosmetic surgery, from reaching audiences between ages 13-17. Moreover, TikTok has invested in creating tools that help people create the best TikTok experience for them. For example:

- People can filter hashtags and keywords from their For You feed;
- They can 'Refresh' their For You feed if their recommendations start to feel stale; and
- They can indicate if they're 'Not interested' in certain content.

5. **If the default settings are different for users aged 16 and 17 than they are for users under the age of 16, please explain why your company takes this position, how this position was developed, and whether any company personnel voiced objections to or raised concerns about this position.**

At TikTok, we work to design tools and policies that promote a safe and age-appropriate experience for teens ages 13-17. While adolescents mature at different rates, based on research and our work with experts we believe that there are generally differences between a 13 year old and 17 year old when it comes to maturity and what may be appropriate, and therefore have even stricter measures to protect younger teens. For example, accounts ages 13-15 cannot have their content Dueted or Stitched, whereas accounts ages 16-17 can restrict who can Duet or Stitch their content.

We also aim to provide parents with resources they can use to have conversations about digital safety and decide the most comfortable experience for their family, including our Family Pairing features[4] and our new Guardian's Guide[5] to TikTok.

6. **What studies, research, summaries, or data does your company have reflecting the efficacy of its parental controls and child safety measures? Please provide these studies, research, summaries, or data.**

Our minor safety team holds a high bar of rigor for developing policy. Our policies are informed by peer-reviewed academic literature and ongoing consultation with external scholars. We work with leading youth-safety and well-being experts, as well as adolescent psychologists, to inform our approach. For example, TikTok partnered with Boston Children's Hospital, which launched a Digital Wellness Lab[6] to serve as a research innovation hub for studying the impact that digital technologies have on the well-being and mental health of children. We seek out feedback, research, and best practices from such experts and organizations, and we use this information to help design TikTok in a way that considers and supports the unique needs of teens. Based on input from these experts and published research in this space, we have adopted a nuanced, age-

[3] https://newsroom.tiktok.com/en-us/more-ways-for-our-community-to-enjoy-what-they-love
[4] https://newsroom.tiktok.com/en-us/supporting-youth-and-families-on-tiktok
[5] https://www.tiktok.com/safety/en-us/guardians-guide/
[6] https://digitalwellnesslab.org/

appropriate approach that distinguishes between early teens (age 13-15) and late teens (age 16-17).

For example, last year every account belonging to a user below age 18 automatically had a 60-minute daily screen time limit. While there's no collectively-endorsed position on the 'right' amount of screen time or even the impact of screen time more broadly, we consulted current academic research and experts from the Digital Wellness Lab in choosing the limit. If the 60-minute limit is reached, teens will be prompted to enter a passcode in order to continue watching, requiring them to make an active decision to extend that time.

Research[7] also showed that being more aware of how we spend our time can help us be more intentional about the decisions we make. So we also decided to prompt teens to set a daily screen time limit if they opt out of the 60-minute default and spend more than 100 minutes on TikTok in a day. This built on a prompt we rolled out the previous year to encourage teens to enable screen time management. In addition, we send every teen account a weekly inbox notification with a recap of their screen time.

7. Concerning international law,

a. what steps have your company and its subsidiaries taken to comply with the European Union's *Digital Services Act*?

In response to the Digital Services Act ("DSA"), we introduced an additional reporting option[8] for our European community (i.e., TikTok users located in EEA countries*)* that will allow people to report content they believe is illegal, including advertising. People can choose from a list of categories such as hate speech, harassment, and financial crimes, and we provided a guide to help people better understand each category.

Under the DSA, we also provide our community in Europe with information about a broader range of content moderation decisions. For example, if we decide a video is ineligible for recommendation[9] because it contains unverified claims about an election that is still unfolding, we let users know. We also share more detail about these decisions, including whether the action was taken by automated technology, and we explain how both content creators and those who file a report can appeal a decision.

We have also been transparent[10] about the recommendation system that powers the For You feed, which sits at the heart of the TikTok experience. As part of our efforts to meet DSA requirements, our European community have another way to discover content on TikTok by turning off personalized recommendations. This means their For You and LIVE feeds will instead show popular videos from both the places where they live and around the world, rather than recommending content to them based on their personal interests. Similarly, when using non-

[7] https://www.internetmatters.org/resources/intentional-use-report/
[8] https://www.tiktok.com/legal/page/global/reporting-illegal-content/en
[9] https://www.tiktok.com/community-guidelines/en/fyf-standards/
[10] https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-foryou

personalized search, they will see results made up of popular content from their region and in their preferred language. Their Following and Friends feeds will continue to show creators they follow, but in chronological order rather than based on the viewer's profile.

Accounts for those aged under 16 are set to private by default globally and their content cannot be recommended in For You feeds. Now, users in Europe aged 13-17 will also no longer see personalized advertising[11] based on their activities on or off TikTok. People already have control over the ads they can see and they can toggle personalized ads on or off in their settings. Finally, we expanded eligibility for our Research API to academic researchers in Europe. These tools were designed to enhance transparency about content on our platform and are informed by feedback we're hearing from researchers and civil society.

More information about these efforts can be found here[12].

b. what steps has your company and its subsidiaries taken to comply with the United Kingdom's *Online Safety Act*?

While the Online Safety Act ("OSA") has been passed by the UK Parliament, it has not substantively come into effect. In particular, much of the practical operation of the OSA is dependent on the completion of a number of consultations by the UK regulator, Ofcom, which are scheduled to take place this year and next. A link to Ofcom's plan for the roll-out of the OSA obligations can be found here[13].

In the meantime, TikTok is regulated under the UK's existing content regulation law, the 'Video Sharing Platform' ("VSP") regulation. TikTok has been complying with this regulation for approximately two years. The VSP includes obligations relating to the protection of minors from content that is not suitable for those under the age of 18 and users more generally from 'harmful material' (e.g., content likely to incite hatred, racism, terrorism or that constitutes the distribution or dissemination of CSAM). An overview of the VSP regulation can be found here[14].

c. what steps has your company and its subsidiaries taken to comply with Australia's *Online Safety Act*?

The Online Safety Act 2021 (Cth) requires all businesses that offer online services to users in Australia to comply with the specific requirements outlined in the Online Safety Act itself, but also the requirements detailed in the Basic Online Safety Expectations ("BOSE"), and the relevant industry code that applies to the business. For TikTok, the relevant industry code is the Social Media Services Code ("SMS Code").

In respect of the requirements under the Online Safety Act itself, TikTok established internal processes to ensure that it can appropriately and efficiently respond to all valid notices issued by

[11] https://www.tiktok.com/business/en/blog/privacy-updates-improved-data-control-transparency-tools
[12] https://newsroom.tiktok.com/en-eu/fulfilling-commitments-dsa-update
[13] https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation
[14] https://www.ofcom.org.uk/online-safety/information-for-industry/vsp-regulation

the eSafety Commissioner. For example, TikTok established a process to ensure that valid takedown notices are actioned and removed within the designated time period (being 24 hours in the case of Class 1 or Class 2 Material, and non-consensual sharing of intimate images, or 48 hours in the case of cyberbullying material targeted towards an Australian child or an Australian adult).

In respect of the requirements under the BOSE, our internal analysis concluded that TikTok already met the 'core expectations' outlined in the BOSE, including but not limited to, the requirement to:

- take reasonable steps to ensure that end-users are able to use the service in a safe manner;
- take reasonable steps to minimize cyber bullying material, non-consensual intimate images of a person, material depicting or promoting abhorrent violent conduct;
- take reasonable steps to ensure measures are in effect to prevent access by children to class 2 material (sexual activity / nudity / drugs / violence);
- ensure that the service has clear and readily identifiable mechanisms to report and make complaints about certain material provided on the service;
- take reasonable steps so that penalties for breaches of terms of use are enforced.

We also took steps to meet the 'additional expectations', including providing the eSafety Commissioner with a designated contact point for online safety matters and adding an additional help center resource for the Australian community. TikTok also responded to its first BOSE reporting notice in 2023, which related to CSEA and CSAM, and eSafety's transparency report (which compares platforms' performance) is available here[15].

In respect of the requirements under the Industry Codes Head Terms and the SMS Code, again our internal analysis concluded that the TikTok platform already met the majority of the requirements under the Code. We also updated our help center resources and introduced a new reporting form for our Australian Community, making it easier for the Community to report any non-compliance with the Code.

d. if those laws create a safer, healthier online experience for kids on your platforms, do you commit to implement these changes in the United States? If not, why not?

TikTok is committed to helping ensure our community has a safe and positive experience on the platform. While this commitment is unwavering, we do not take a one-size-fits all approach, and a regional focus allows us to better understand local challenges and trends and develop informed solutions that consider unique local context and cultures. This commitment to localization enables us to create a more targeted, responsive approach to safety and enables us to stay up-to-date with the latest developments in each region.

[15] https://www.esafety.gov.au/sites/default/files/2023-10/Full-transparency-report-October-2023.pdf

We work with our Regional Safety Advisory Councils and U.S. Content Advisory Council to bring together groups of independent experts who help us develop forward-looking policies and processes that not only address the challenges of today, but also plan ahead for the next set of issues that our industry will face. These councils are an important way to bring outside perspectives into our company and onto our platform.

Our council members represent a diverse array of backgrounds and perspectives, and are made up of experts in youth safety, free expression, hate speech, and other safety topics. They work collaboratively with us to inform and strengthen our policies, product features, and safety processes.

We have established eight regional Safety Advisory Councils in Asia Pacific, Brazil, Europe, Latin America, MENAT (Middle East, North Africa, Turkey), and a U.S. Content Advisory Council. We aim to continue expanding our regional presence.

8. **Child predators have been found to use TikTok's "Only Me" privacy setting to lure children into sharing CSAM with them. This setting allows users to upload private videos that are only visible to the user. Child predators exploit this feature by sharing a password with children and then asking those children to upload CSAM to the shared account using "Only Me."**

   What steps does TikTok take to ensure "Only Me" isn't misused in this way?

We disagree with this assertion. When we find accounts attempting to obtain or distribute CSAM, we remove them, ban their device so they cannot create another TikTok account on the same device, and make reports to NCMEC. Whenever a video is uploaded to the TikTok platform -- including posted videos that are viewable only to the account holder -- the content is run through our automated moderation process. During this process, our systems work to detect and remove violations of our Community Guidelines, including CSAM, or flag content for human evaluation. If moderators find violations of our policies, the content will be removed. We provide special guidance to moderators to help ensure they are alert to signs of such behavior or content.

9. **Although TikTok bans many hashtags associated with CSAM, by making slight variations in the spelling of CSAM hashtags, predators are able to easily locate victims and CSAM on the platform. Additionally, TikTok's algorithm at times actually promotes these accounts to users who have demonstrated interest in similar accounts.**

   In addition to banning hashtags, what other steps does TikTok take to combat CSAM and prevent its algorithm from actively promoting these accounts?

When TikTok bans a keyword, we also ban variations including misspellings and will ban any new variations we surface. Additionally, TikTok implements logic which identifies users displaying risky behaviors on platform (such as using problematic keywords, being blocked by minors, or posting or sharing sexualized content related to or featuring minors) and restricts

recommendations for these accounts. Accounts that exhibit potentially predatory behavior (i.e., where we do not have enough signal of a violation of our Community Guidelines) are not suggested to any users, and we also reduce exposure between such accounts and minors.

10. **In September 2021, TikTok announced it had 1 billion monthly users. In 2022, TikTok sent 288,125 reports to NCMEC's CyberTipline. When comparing these numbers to other platforms, there is a clear discrepancy in the ratio of users to reports made. For example, Discord had only 150 million monthly users but made 169,800 reports in 2022.**

    How does TikTok explain the discrepancy in the number of reports it submitted to NCMEC?

TikTok's overall reports to NCMEC doubled between 2019 and 2022. However, we have no indication that increased numbers year-over-year mean that CSAM went undetected in prior years. For instance, NCMEC has a process for reaching out to platforms when it detects potential CSAM before the platforms do. The number of requests TikTok has received from NCMEC is extremely small compared to peer platforms, and we have addressed those requests promptly. To respond directly to your example, in 2022, TikTok received only one request from NCMEC[16].

Each platform operates in very different ways, making direct comparisons difficult. For our part, TikTok makes deliberate design choices that make our platform inhospitable to those would seek to find or share child sexual abuse material. We realize that there is no finish line when it comes to keeping children safe, and we will always work to improve. To that end, TikTok is constantly working to evolve our detection capabilities. The more tools we bring to bear, the more effective we become at detecting and removing this content.

There is also a difference in how different platforms are designed. For example, there are no private groups on TikTok, and TikTok's direct messaging system is not conducive to sharing violative content. For instance, accounts registered to users under 16 cannot send or receive direct messages, and users 16 or over can only send direct messages to mutual followers.

11. **TikTok claims that it has a zero-tolerance policy for child sexual abuse material. Yet, in a 2022 Forbes article, TikTok's livestreaming function was described by one expert as "the digital equivalent of going down the street to a strip club filled with 15-year-olds." TikTok Live allows users to livestream real-time videos. Other users can view the livestreams and send digital gifts to the host, which can then be redeemed for real money.**

    Despite TikTok's policies prohibiting users under the age of 18 from hosting livestreams and sending or receiving digital gifts, underage users are being victimized and exploited on TikTok Live. Minors as young as fourteen are hosting livestreams for audiences of child predators, who use the comment function to coerce these children into performing explicit acts in exchange for digital gifts.

[16] https://www.missingkids.org/content/dam/missingkids/pdfs/2022-notifications-by-ncmec-per-esp.pdf

Viewers often record this content and generate CSAM that is spread across third-party websites.

What steps is TikTok taking to combat this problem?

TikTok account holders must be at least 18 years of age to livestream, and must also have a certain number of followers. TikTok's zero-tolerance policy on CSAM applies to livestreams. Livestreams are regularly monitored through automated processes and human moderation to detect violative content, including CSAM. Where appropriate, livestreams are interrupted and taken down. Any livestream or livestream comment that is detected and taken down that involves CSEA is also reported to NCMEC.

We use a variety of methods to detect CSAM on livestreams, but do not publicly disclose those methods in order to prevent bad actors from attempting to circumvent our systems. We would be willing to privately brief the Committee on these methods.

TikTok's Responses to Questions for the Record from Senator Graham
Submitted March 7, 2024

1. **Do you support S. 1207, the bipartisan EARN IT Act? Why or why not?**

We support efforts to address the challenges of child sexual abuse material (**"**CSAM") and are open to working closely with Congress for legislative solutions that address the problem directly. However, as currently drafted, the EARN IT Act would expand liability for platforms that make substantial efforts to eliminate CSAM. We are also concerned that it may lead to unintended consequences, such as overmoderation and censorship. We believe that legislation, such as the Invest in Child Safety Act, can effectively address the challenge of CSAM by significantly increasing resources for law enforcement agencies to investigate and prosecute CSAM reports made to the National Center for Missing and Exploited Children ("NCMEC").

2. **What measures are you taking to prevent and address sextortion, including financial sextortion, on your companies' platforms?**
    a. **What methods are in place to detect and disrupt this type of abuse in real time?**

We do not allow youth exploitation and abuse, including nudity, grooming, sextortion, solicitation, pedophilia, and physical or psychological abuse of young people. Our real time detection efforts can be seen in our most recent Community Guidelines Enforcement Report[1], which shows that we proactively removed roughly more than 96% of identified violative content before it was reported.

We want everyone to feel safe and comfortable when they spend time on TikTok. That's why our Community Guidelines apply to everyone and everything on the platform, whether videos, livestreams, or comments. TikTok has always had strict policies prohibiting nudity, sexual activity, and sexually explicit content, including content that directs users to adult websites or apps. This includes content categories that are not eligible for recommendation, including implied nudity, sensual content, and other content that may be allowed on the platform, but is not suitable for all audiences over age 13.

TikTok reports to NCMEC text-based violations both for comments and direct messages. These reports usually involve cases of online enticement behavior, but can also include grooming, minor sexual solicitations and sextortion cases.

TikTok also participates in NCMEC's Take It Down[2] service, which is a free service that can help process requests to remove or stop the online sharing of nude, partially nude, or sexually explicit images or videos taken when someone is under 18 years old.

[1] https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement-2023-2/
[2] https://takeitdown.ncmec.org/

3. Please provide the committee statistics on how long it takes your company to respond to various types of legal process from law enforcement?

In 2023, on average, TikTok responded to legal process requests within 29 days.

Type of Legal Process	Average (Days)
Preservation/Preservation Extension Requests:	9.8
Subpoenas/Grand Jury Subpoenas:	23.8
Court Order	12.1
Search Warrant	37.9
PR/TT	42.8
Total Average	29.2

4. Do you notify your users when law enforcement serves subpoenas/summons for subscriber information and specifically requests not to notify the subscriber/user?

a. If you notify the subscriber, how long do you wait until notification goes out?

TikTok currently does not notify U.S. users when served with legal process by law enforcement. Should we provide user notice in the future, we will comply with valid non-disclosure orders. The current process for non-government and civil legal requests ensures the notification is sent within 3 days of intake. Upon Notice, TikTok provides a deadline of 14 calendar days (unless the legal request provides alternate specific direction).

b. Are you aware that by notifying the subscriber about a law enforcement subpoena for their subscriber information that you are jeopardizing critical evidence that could be erased before law enforcement can serve warrants?

We do not currently notify U.S. users when law enforcement serves legal process. Should we provide user notice in the future, we will comply with valid non-disclosure orders.

c. Would your company agree to a 90-day non-disclosure to subscribers to allow law enforcement ample time to secure proper legal process?

We do not currently notify U.S. users when law enforcement serves legal process. Should we provide user notice in the future, we will comply with valid non-disclosure orders.

5. **Do you actively seek out and incorporate feedback and insight from survivors of online sexual exploitation to improve your trust and safety policies and practices and to prevent and disrupt child sexual abuse material (CSAM) production and distribution on your platform? Can you provide examples?**
    a. **If not, please explain.**

We engage closely with organizations that work directly with survivors of online sexual exploitation, who facilitate these conversations and provide insights. These groups include:

- Internet Watch Foundation[3]
- WePROTECT Global Alliance[4]
- National Center for Missing & Exploited Children[5]
- International Justice Mission[6]
- National Center on Sexual Exploitation[7]

6. **During our hearing, you testified that you collaborate with parents and parent organizations to create mechanisms to keep children safe online. Please elaborate and cite examples of your company's work with non-employee parents and parent organizations.**

TikTok is continuously working to provide a safe app experience for our community, and we aim to be a leader in this area. We recognize, however, that technology is ever-evolving and that we need to be prepared to address unexpected trends and challenges as they arise. Beyond our efforts with NCMEC, TikTok regularly consults with and listens to organizations that support minor safety efforts:

- We worked with ConnectSafely—a nonprofit dedicated to educating users of connected technology about safety, privacy, and security—to develop a TikTok-specific guide for parents and teens.
- TikTok is a member of Family Online Safety Institute ("FOSI"), which convenes leaders in industry, government, and the non-profit sectors to collaborate and innovate new solutions and policies in the field of online safety. We've partnered with FOSI on TikTok's Top 10 Tips for Families guide and to create the TikTok Tools & Resources For Families guide[8] to help parents and teens create a fun, safe, and age-appropriate experience on the platform. Among other things, this guide provides steps to report content if a user believes it violates our Community Guidelines, and links to additional resources such as our Youth Portal and Guardian's Guide.

[3] https://www.iwf.org.uk/
[4] https://newsroom.tiktok.com/en-us/tiktok-joins-weprotect-global-alliance
[5] https://www.missingkids.org/HOME
[6] https://www.ijm.org/our-work
[7] https://endsexualexploitation.org/about/
[8] https://www.fosi.org/good-digital-parenting-tool/tiktok-resources-for-families

- We are an active member of the Tech Coalition, a global alliance of technology companies to protect children online. In addition to being on its board, we also co-led the multi-stakeholder forum on Minor Financial Sextortion held last year, and we chair committees within the organization to advance the fight against online child sexual exploitation and abuse.
- We are part of the WePROTECT Global Alliance, the largest and most diverse multi-sector alliance dedicated to ending online child sexual exploitation.
- Internet Watch Foundation ("IWF") is a vital partner for TikTok in our work to counter online child sexual abuse and exploitation. TikTok accesses IWF's URL and keyword list and, via NCMEC, its hash database of known child sexual abuse images. In addition to its frontline work, IWF provides insight on new and emerging trends and acts as a convener for key stakeholders.
- TikTok partners with the Boston Children's Hospital's Digital Wellness Lab, to provide resources for families, educators, and clinicians and research to understand and promote child and family wellness.
- In addition, we host global Advisory Councils[9], and have formally established a Youth Council[10] that directly engages with teens to solicit their input.

Additionally, TikTok is part of the alliance in support of the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse. This alliance brings together governments across the world, including the United States; digital platforms and other private companies; and civil society and intergovernmental organizations. The objective is to collaborate across sectors and industries to address online threats to minors. TikTok also works with a group of non-governmental organizations as part of a CSAM intervention project designed in part to help direct users to appropriate resources.

Beyond the partnerships with leading organizations in the space, we see others finding great success in leveraging TikTok as a platform to spread critical safety messages. For example, Thorn launched an in-app campaign called NoFltr, which facilitated a conversation between youth and adults on the sharing of nude images and consent. We also worked in collaboration with Thorn researchers and our policy team to inform our policy and features related to their findings.

We engage in these partnerships because online threats are complex and dynamic. No single company or government can solve these problems in a vacuum. We look forward to continuing our own work, as well as our collaborations, to make a meaningful difference in keeping our community safe.

[9] https://www.tiktok.com/transparency/en/advisory-councils/
[10] https://newsroom.tiktok.com/en-us/updating-family-pairing-and-establishing-tiktoks-youth-council

7. **Why does your company have the age limit of 13 years old for a user to sign up for an account?**
    a. **Why not younger or older?**

TikTok is deeply committed to ensuring that its platform provides a safe and positive experience, especially for people under the age of 18. TikTok follows the Federal Trade Commission's ("FTC") guidance and provides a separate experience in the U.S. for users under 13. This curated viewing experience has additional safeguards and privacy protections designed specifically for a younger audience. In line with the FTC's guidance on children's privacy, users 13 and older may access the full TikTok experience.

TikTok provides additional safeguards for teen users, including: (1) a default daily screen time limit of 60 minutes, (2) setting accounts to private by default, (3) restricting access to host LIVE content or engage in financial transactions, and (4) turning off push notifications at night. Users under 16 may not have content featured in the For You feed, may not send or receive direct messages, may only receive comments on their content from friends, and their videos are not available for duets or stitches. These measures help further TikTok's goal of providing young people with an experience that is developmentally appropriate and help ensure a safe space for self-exploration.

8. **How many minors use your platform? How much money does your company make annually from these minors?**

TikTok has an average 170 million monthly average users and we believe that the average U.S. user is over 30 years of age. As a privately held company, TikTok does not disclose detailed user numbers or revenue.

9. **What percentage of your employees work on trust and safety and how much money does your company invest annually in trust and safety?**

TikTok currently has more than 40,000 trust and safety professionals working to protect our community. We expect to invest more than two billion dollars in trust and safety efforts in 2024, with a significant part of that investment in our U.S. operations.

10. **It is sometimes challenging for law enforcement conducting criminal investigations to determine the true identity of a person behind a name on social media or other online platforms, and whether an online identity is an actual person. What are you doing to validate the true identity of users – or the fact that a user is a human – when they create an account on your platforms?**

Like our industry peers, TikTok does not verify user identity for most accounts. We do verify user identity in certain circumstances including the verification badge process for high profile accounts and the process that allows us to make payments to certain creators. A verified badge means that TikTok has confirmed the account belongs to the person or brand it represents. In order to verify an account, users submit a verification request, and TikTok collects information to help ensure that the account holder can be verified.

To build a trusted community online, our Community Guidelines prohibit account behaviors that may spam or mislead our community. The following behaviors are not allowed on TikTok:

- Spam**,** including
    - Accounts that are operated: (1) in bulk, (2) through unauthorized automation, or (3) in order to distribute high-volume commercial content
    - Operating networks of accounts that represent similar entities or post similar content to lead others to specific locations (on or off-platform), such as other accounts, websites, and businesses
- Impersonation, including:
    - Accounts that pose as another real person or entity, such as using someone's name, biographical details, content, or image without disclosing it
    - Presenting as a person or entity that does not exist (a fake persona) with a demonstrated intent to mislead others on the platform

If we determine someone has engaged in deceptive account behaviors such as spam or impersonation, we will ban the account[11], and may ban any new accounts that are created.

11. Is your company using safety technology to detect and prevent live video child sexual abuse on your platforms and apps that allow users to stream or share live video? If not, please explain.

Yes. Internally, we have developed a wide range of technology capabilities to detect both content risk (e.g., LIVE video depiction of child abuse, CSAM content, text based content) and behavior risk (e.g., predator, groomer). We do not publicly disclose those methods in order to prevent bad actors from attempting to circumvent our systems, but would be willing to privately brief the Committee on these methods.

a. Has your company tested that or similar technology? If not, are you developing similar technology to address child sexual abuse in live video?

Yes. For example, in 2023 we conducted an evaluation with a trusted third party vendor sponsored by the Technology Coalition, and our internal LIVE detection methods outperformed the vendor technology.

[11] https://www.tiktok.com/community-guidelines/en/accounts-features?cgversion=2023

12. How are you measuring if your trust and safety policies, practices, and tools are effective in protecting children from sexual abuse and exploitation on your platform?

To ensure our policies stay up-to-date and effectively capture emerging risks and trends, TikTok regularly engages with industry, NGOs, academics, civil society, and other relevant organizations, including NCMEC and those referenced in our response to your Question 5.

We also believe that listening to the experience of teens is one of the most important steps we can take to build a safe platform for teens and their families. It helps us avoid designing teen safety solutions that may be ineffective or inadequate for the actual community they're meant to protect, and it brings us closer to being a strong partner to caregivers as we can better represent teens' safety and well-being needs. We launched TikTok's global Youth Council, where we listen to the experiences of those who directly use our platform and be better positioned to make changes to create a safe experience for our community.

a. What specific metrics or key performance indicators do you use?

We review the efficacy of our policies, detection and enforcement based on several metrics, including the volume, accuracy and speed of removal of content that violates our Community Guidelines, the prevalence of CSAM and Child Sexual Exploitation and Abuse ("CSEA") detected at creation and before any interactions or engagements, as well as the number of cases we report to NCMEC. Our quarterly Community Guidelines Enforcement Reports[12] also provides insight into our enforcement efforts, showing how we continue to uphold trust, authenticity, and accountability.

13. Is your company using language analysis tools to detect grooming activities? If not, please explain.

Yes, TikTok uses various language analysis tools to detect grooming activities, such as keyword lists, Natural Language Processing, and URL detection. We also work with the Internet Watch Foundation to continuously update CSAM and grooming related keywords and URL from the industry.

a. What investments will your company make to develop new or improve existing tools?

TikTok is investing in language analysis and other technical tools to understand evolving predator behaviors across all regions, including the U.S. We are building models to detect grooming and predator behaviors across the TikTok platform, as well as feature-specific models trained to better detect unique behaviors within a particular feature (e.g., DM, Live). We expect to launch these models in H2 2024. We also continuously refresh our keywords with new terms to reflect evolving behaviors and develop new strategies to help reduce the risk of predatory interactions.

[12] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/

14. What resources have you developed for victims and survivors of abuse on your platforms?

TikTok has developed a safety center for survivors of sexual abuse[13] that provides resources for help and information in more than 30 countries. The resources can be accessed either by going directly to the safety center or by search terms/keywords associated with abuse through an in-app response that redirect our community members to supportive resources. The pages also links to StopNCII.org[14], which allows people to report non-consensual sexual images (also referred to as image-based sexual abuse). StopNCII translates those images to hashes and shares them with TikTok and other companies to remove them from the app. TikTok also participates in NCMEC's Take It Down service.

15. What voluntary hash-sharing or other information sharing initiatives does TikTok participate in to help combat child sexual exploitation?

TikTok participates in IWF hash, URL, and Keyword lists, NCMEC NGO and Industry Hash list, Take It Down, and has applied to join the Tech Coalition's new Lantern project. We have also integrated Google's Content Safety API to further support the proactive identification of never-before-seen CSAM imagery, as well as YouTube's CSAI Match to further support detection of known CSAM in videos.

16. What safety messaging does TikTok provide to its younger users around online safety, especially as it relates to online enticement and financial sextortion?

TikTok provides messaging on sexual assault and CSEA prevention and resources on its Safety Center[15] and Youth Portal[16]. The Youth Portal specifically highlights how to keep your account secure, manage privacy and safety settings, and restrict unwanted interactions. Consistent with enforcement of our Community Guidelines, TikTok also blocks certain messages from being sent within Direct Messaging and sends a risk notice to users with instructions on how to report the message.

17. You testified that TikTok plans to invest $2 billion in trust and safety globally. How much will be spent on trust and safety in the United States?

We expect to invest more than two billion dollars in trust and safety efforts in 2024, with a significant part of that investment in our U.S. operations.

[13] https://www.tiktok.com/safety/en/sexual-assault-resources/
[14] http://stopncii.org/
[15] https://www.tiktok.com/safety/en-us/sexual-assault-resources/
[16] https://www.tiktok.com/safety/youth-portal?lang=en

18. You further testified that TikTok is not in a position to share financials publicly. Please provide the committee with how much revenue TikTok generated in the past three years.

As a privately held company, TikTok does not disclose revenue.

19. What investigative steps have been taken since a senior TikTok officer, Barak Herscowitz, in the Israel office resigned due to other TikTok employees celebrating the barbaric acts of Hamas and other Iranian backed terror groups?

a. Your testimony states that: "Investing in teams and technology is a core priority for me as CEO." As CEO, have you directed that those Trust & Safety officers jubilantly cheering for Hamas terrorists be fired?

b. Will you provide my office with a copy of Mr. Herscowitz's internal memo regarding TikTok moderators openly expressing support for terrorism and actively promoting Hamas?

c. Will you provide my office with copies of Lark communications that show TikTok moderators' and Trust & Safety Officers' expressions of support for terrorist groups, including Hamas, as well as expressions of antipathy to Israel?

This report is inaccurate in several respects. Mr. Herscowitz worked for TikTok in a sales role; he therefore was not involved in content policy, nor was he part of the process for ensuring employees abide by our internal rules in the workplace.

TikTok has strong policies against discrimination and harassment in the workplace, and praise in the workplace for the October 7th attack or for any other forms of terrorism would violate these policies. Employees are encouraged to report their concerns, anonymously if they so choose, and every incident is investigated by the appropriate internal team. Documents or communications relating to any such investigations are confidential.

TikTok's priority is to keep both its global community on the platform and those impacted by these tragic events safe. This sentiment was reiterated in a message sent to all TikTok employees denouncing the October 7th attacks, as well as in our published statement online[17].

20. Will you fully restore the Creative Center tool so that TikTok's claims of independence from Beijing can be substantiated? If not, please explain.

This resource is designed to provide brands with the top trending content to help them better understand trends. Unfortunately, some individuals and organizations have misused the Center's search function to draw inaccurate conclusions, so we changed some of the features to ensure it is used for its intended purpose. To continue serving advertisers, we allow searches of hashtags that are in the top 100 by industry. For content research purposes, we provide a Research API that enables academic researchers to independently study content.

[17] https://newsroom.tiktok.com/en-us/our-continued-actions-to-protect-the-tiktok-community-during-the-israelhamas-war

TikTok's Responses to Questions for the Record from Senator Whitehouse

1. **What exemptions from the protections of Section 230 would your company be willing to accept?**

TikTok welcomes the opportunity for conversation about ways that Section 230 could be further updated. The Communications Decency Act of 1996 generally immunizes interactive computer services from liability as a publisher or speaker for online content posted by their users. But the statute does not apply to—and, indeed, specifically excludes—claims related to certain federal obscenity statutes, federal criminal statutes, and types of intellectual property claims. In 2018, Congress enacted an additional exception to Section 230 immunity for certain claims related to sex trafficking. Enforcing laws against, and taking action to curtail, sex trafficking and child sexual exploitation crimes, are critical to protecting children. Any updates to Section 230 should seek to balance all of Section 230's important policy goals, including in promoting the continued development of the internet as a "vibrant and competitive" market, while also taking measures to help ensure that users are safeguarded against criminal acts and exploitation.

2. **Is it your belief that your company should enjoy absolute immunity under Section 230 from suits like Doe v. Twitter, No. 21-CV-00485-JCS, 2023 WL 8568911 (N.D. Cal. Dec. 11, 2023), no matter the extent of your company's failure to remove reported child sexual abuse material from the platform or to stop its distribution?**

Section 230 of the Communications Decency Act of 1996 generally immunizes interactive computer services from liability for online content posted by their users, but its protections are not absolute. Congress has already created exceptions to Section 230 immunity, including for certain claims related to certain federal obscenity statutes and related to sex trafficking.

TikTok's Responses to Questions for the Record from Senator Coons
Submitted March 7, 2024

1. **Does TikTok Inc. ("TikTok") measure an estimated total amount of content on the platform that violates its suicide and self-harm policy? If not, why not?**

As explained in more detail in response to Question 1(a), TikTok measures the volume of videos removed for violating our suicide and self-harm policies.

   a. **Does TikTok disclose an estimated total amount of content on its platform that violates its suicide and self-harm policy? If so, please provide a specific citation to where TikTok discloses that information. If not, why not?**

TikTok publishes Community Guidelines Enforcement Reports[1] each quarter. In its latest report, covering Q3 2023 we shared that 0.8% of videos published during the time period were removed (136M total videos). Of the videos removed, 10.4% were removed for violating our policies for mental and behavioral health. Of the videos removed under our mental and behavioral health policy, 23.4% were removed for violations related to suicide and self-harm. Accordingly, 0.02% of videos uploaded to TikTok were removed for violating our policies against Suicide and Self-Harm[2]. Additionally, of this content, TikTok proactively identified and removed 97.9% before it is reported to us; more than 90% of this content was removed within 24 hours of posting.

Unlike our competitors, TikTok publishes these statistics for the U.S. (and 49 additional markets) as well as the global aggregates.

2. **TikTok has previously reported how much content it removes under the platform's suicide and self-harm policy.**

   a. **For content that has been removed, does TikTok measure how many views that content received prior to being removed? If not, why not?**

We do track the percentage of content removed with zero views.

   b. **For content that has been removed, does TikTok disclose how many views that content received prior to being removed? If so, please provide a specific citation to where TikTok discloses that information. If not, why not?**

We disclose the percentage of content removed before receiving any views in our Community Guidelines Enforcement Reports[3] each quarter.

[1] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/
[2] https://www.tiktok.com/community-guidelines/en/mental-behavioral-health/#1
[3] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/

c. Please provide an estimate of the number of views content that was removed under this policy received in January 2024.

In Q3 2023, over 75% of content identified and removed for violating our policy on Suicide and Self-Harm was removed before receiving any views. In the same time period, views on content identified and removed for violating this policy represented approximately 0.00017% of total views on the platform.

d. For content that has been removed, does TikTok measure demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If not, why not?

We do not track viewer demographics. TikTok prioritizes tracking removals based on the potential severity of policies they violate (i.e., the basis for the content's removal), and not viewer demographics.

e. For content that has been removed, does TikTok disclose demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If so, please provide a specific citation to where TikTok discloses that information. If not, why not?

As mentioned in our response to 2(d), we do not track this information and therefore do not disclose it.

f. Does TikTok measure the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If not, why not?

For content that is ultimately removed, we do not track whether it was viewed by a user multiple times prior to its removal.

TikTok's policies aim to support people who may be struggling, and we have developed in-app interventions in furtherance of this effort. For example, we will not show self-harm related content when searching related terms. Instead, we want to support our community by providing resources and access to emotional support helplines. We surface regional suicide prevention hotline numbers and additional localized resources that can help. Additionally, users cannot create LIVE rooms with suicide or self-harm keywords in title and cannot post a video or a LIVE that was removed for violating our suicide or self-harm policies.

g. Does TikTok disclose the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If so, please provide a specific citation to where TikTok discloses that information. If not, why not?

As mentioned in our response to 2(f), we do not track this information and therefore do not disclose it.

3. **TikTok utilizes an algorithm to recommend or amplify content to users.**

    a. **For content that has been removed, does TikTok measure whether and the extent to which the removed content was recommended or amplified by TikTok? If not, why not?**

We strive to remove violative content as soon as possible, and before it's reported to us by our community. In our latest Community Guidelines Enforcement Report[4] covering Q3 2023, the removal rate for violative content was 96.1% before a single view and 90.6% within 24 hours of posting. TikTok does not measure whether violative content was amplified as we have separate teams that are responsible for identifying and removing violative content, and other teams responsible for recommending content.

    b. **For content that has been removed, does TikTok disclose whether and the extent to which the removed content was recommended or amplified by TikTok? If so, please provide a specific citation to where TikTok discloses that information. If not, why not?**

TikTok does not measure or disclose this.

    c. **For content that has been removed, does TikTok measure how many views the removed content received after having been recommended or amplified? If not, why not?**

We measure and include in our Community Guidelines Enforcement Report[5] the proportion of content that is removed without views.

    d. **For content that has been removed, does TikTok disclose the number of views the removed content received after having been amplified or recommended? If so, please provide a specific citation to where TikTok discloses that information. If not, why not?**

Any content that violates TikTok's Community Guidelines we aim to remove as swiftly as possible and limit the exposure of views on this content. We disclose the proportion of content that is removed without a single view. In our latest Community Guidelines Enforcement Report[6] covering Q3 2023, 76.8% of the videos removed had 0 views. Less than 1% of videos published on TikTok are ever removed for violating our Community Guidelines.

4. **Does TikTok support creating industry-wide transparency requirements to disclose basic safety information, like those included in the Platform Accountability and Transparency Act?**

[4] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/
[5] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/
[6] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/

TikTok supports and has been a leader in platform transparency. TikTok strives to foster a fun and inclusive environment where people can create, find community, and be entertained. To maintain that environment, we take action upon content and accounts that violate our Community Guidelines[7] or Terms of Service[8] and regularly publish information about these actions to hold ourselves accountable to our community. We release quarterly reports[9] to bring transparency to the actions we take to help keep TikTok safe, welcoming, and entertaining for our global community. As part of our continued efforts to make it easy to study the TikTok Platform, we continue to expand our reporting. For example, as of our most recent report (published on December 13, 2023), we now provide additional removal data by policy category for the 50 markets with the highest volumes of removed content in the downloadable data file below. These markets account for approximately 90% of all content removals for this quarter. As we continue our work to build a safe, inclusive and authentic home for our global community, we look forward to sharing more on our evolving efforts to prevent harm.

In the spirit of PATA, TikTok also supports independent research through our Research API. Using the Research API, non-profit universities in the U.S. and Europe can apply to study public data about TikTok content and accounts. We're working to provide increased access to the Research API in the future. Since we introduced the Research API, we have rolled out improvements based on feedback we've heard from the community. We're dedicated to hearing and incorporating feedback from testers and creating an API that will meet the needs of scientific researchers while respecting the privacy of our community.

[7] https://www.tiktok.com/community-guidelines/en/
[8] https://www.tiktok.com/legal/page/us/terms-of-service/en
[9] https://www.tiktok.com/transparency/en/reports/

TikTok's Responses to Questions for the Record from Senator Booker
Submitted March 7, 2024

1. **Trust and safety teams are a vital component in combatting the spread of CSAM, hate speech, violence, and other violative content on tech platforms. Despite this, tech companies have time and time again disinvested from their trust and safety team, especially during changes in leadership.**

    a. **How has the size of your trust and safety team changed over the past five years? Please provide numbers for each of the past five years.**

TikTok currently has more than 40,000 trust and safety professionals working to protect our community, and we expect to invest more than two billion dollars in trust and safety efforts this year alone, with a significant part of that investment in our U.S. operations. We are continuing to hire for our Trust and Safety team globally, with 1,000+ open roles.

    b. **Do your trust and safety teams make submissions to the National Center for Missing & Exploited Children's CyberTipline, or is that a separate unit?**

The Child Safety Team, which is a part of our Trust and Safety team, makes submissions to NCMEC.

    c. **If it is a separate unit, how many members are on the team and how have those numbers changed over the past five years. Please provide numbers for each of the past five years.**

As noted in our response to Question 1(b), the Child Safety Team is part of the Trust and Safety team.

2. **The National Center for Missing & Exploited Children's CyberTipline plays an integral role in combatting child sexual exploitation. The tipline helps law enforcement investigate potential cases and allows prosecutors to bring justice to victims. While federal law requires your company to report to the CyberTipline any apparent violations of federal laws prohibiting child sexual abuse material of which you are aware, there are many gaps.**

    a. **Is there a standard format your reports to the CyberTipline follow? If so, what is that format?**

TikTok utilizes two methods for submitting reports to NCMEC. Our child safety teams utilize both direct reporting through an API (developed with the support of NCMEC) and the manual NCMEC provided webform for CyberTipline reports. These two formats disclose similar information with slight variations due to the technical differences in the reporting method. TikTok produces: the identified user content depicting CSAM, metadata associated with the

CSAM content, subscriber information. TikTok also produces IP data: through the API, we provide the three most recent IP addresses; through manual reports, we produce the data from the last month.

b. Does your company proactively report planned or imminent offenses?

Yes, TikTok reports exigent circumstance cases to NCMEC via their ESP escalation button that prioritizes these reports. Additionally, TikTok also coordinates with its Law Enforcement Outreach team for emergency cases globally, who will directly engage directly with local law enforcement. This is especially pertinent as the CyberTipline report number helps fast track a review and response.

c. Does your company proactively report potential offenses involving coercion or enticement of children?

Yes, TikTok reports text-based violations both for comments and direct messages ("DMs"). These reports usually involve cases of online enticement behavior, but can also include grooming, minor sexual solicitations and sextortion cases.

d. Does your company proactively report apparent child sex trafficking?

Yes, TikTok reports cases of suspected child sex trafficking.

TikTok's Responses to Questions for the Record from Senator Padilla
Submitted March 7, 2024

1. **In recent years, more companies in the tech sector are offering tools to enable caregivers to have a dialogue with minors in their care about healthy and safe internet activity. An important element in understanding whether these tools are helpful is understanding whether or not these tools are being adopted.**
    a. **How many minors are on TikTok?**
    b. **Of these minors, how many of them have caregivers that have adopted Family Pairing?**
    c. **How are you ensuring that young people and their caregivers are aware of these tools?**
    d. **How are you ensuring that these tools are helpful to both minors and their caregivers?**

TikTok has an average 170 million monthly average users and we believe that the average U.S. user is over 30 years of age. As a privately held company, TikTok does not disclose detailed statistics about its users or adoption of specific features.

TikTok has invested significant resources in promoting Family Pairing, including utilizing both earned and paid media such as advertisements, billboards, and local news. Family Pairing has also been promoted to users via in-app notifications and on our Guardian's Guide[1].

In addition, we host global Advisory Councils[2], and have formally established a Youth Council[3] that directly engages with teens to solicit their input.

We also regularly work with experts[4] in online security, wellness, digital literacy, and family safety to help provide advice and resources for our community.

2. **TikTok offers a broad range of "user empowerment" tools, and it's helpful for policymakers to understand whether young people even find these tools helpful or are actually adopting them. Additionally, some safety features still put the onus on young people to employ a great deal of judgment about safety.**
    a. **Last year, TikTok rolled out changes to help teens manage their time on TikTok. What impact has this feature had on the amount of time young people spend on TikTok?**
    b. **How are you ensuring that the burden is not on young people to make adult-level decisions about safety on TikTok?**

[1] https://www.tiktok.com/safety/en/guardians-guide/
[2] https://www.tiktok.com/transparency/en/advisory-councils/
[3] https://newsroom.tiktok.com/en-us/updating-family-pairing-and-establishing-tiktoks-youth-council
[4] https://www.tiktok.com/safety/en/safety-partners/

c. Over the last 4 years, how often have you blocked products from launching because they were not safe enough for minors, or withdrawn products from the market after receiving feedback on the harms they were causing?

TikTok is committed to the principles of safety by design and privacy by design. When features do not meet these standards, they will not be approved for launch.

We help our community understand and control how they spend their time on TikTok. For example, we offer:

- Screen time dashboards that provide insight into how and when a community member is using TikTok;
- Screen time breaks that nudge our community members to take a break from the app after a period of uninterrupted screen time;
- Sleep reminders that allow people to set a reminder to log off at a certain time of day;
- Daily screen time limits that allow people to determine on how much time they spend on TikTok each day; and
- Screen time updates that allow people to receive weekly info about their screen time usage.

TikTok promotes a safe and age-appropriate experience for teen users between ages 13 through 17 by utilizing a multi-faceted, multi-level approach that currently includes the following age-based default settings:

- Accounts ages 13-15:
  - Have a 60 minute screentime limit by default
  - Cannot send or receive virtual gifts[5]
  - Are set to private by default
  - Cannot access direct messaging
  - Cannot host a livestream
  - Cannot buy or sell on TikTok Shop
  - Are not able to have their content recommended to people they do not know in the For You feed
  - Cannot have their content Dueted or Stitched
  - Do not receive push notifications from 9 pm to 8 am
  - Do not have their accounts recommended to others by default. If the user changes this option, their account still will not be recommended to people 18 and over

[5] https://newsroom.tiktok.com/en-gb/updating-our-gifting-policies

- Accounts ages 16-17:
  - Have a 60 minute screentime limit by default
  - Cannot send or receive virtual gifts
  - Private account option is pre-selected by default at account registration
  - Cannot host a livestream[6]
  - Cannot buy or sell on TikTok Shop
  - Do not receive push notifications from 10 pm to 8 am
  - Do not have their accounts recommended to others by default. If the user changes this option, their account still will not be recommended to people 18 and over

In addition, we offer parental controls. Family Pairing allows a parent or guardian to link their TikTok account to their teen's account to directly manage a number of safety controls for their teen's account, including:

- Account privacy: Set their teen's account to private or public.
- Comments: Restrict who can comment on their teen's videos.
- Direct Messages: Restrict who, if anyone, can send private messages to the account
- Daily screen limits: Set the amount of time spent on TikTok each day
- Screentime dashboard: See their teen's time spent on TikTok
- Search settings: Restrict their ability to search for content
- Push notifications: Restrict push notifications to their teens during certain times
- Keyword filters: Add hashtags or keywords they would prefer their teen not see in their For You feed recommendations
- STEM feed: Ensure TikTok's STEM feed - a feed featuring videos related to science, technology, engineering, and math - is enabled on their teen's account

3. **Existing detection tools for keeping child sexual abuse material from spreading online rely on hashed images of already identified CSAM imagery. There are tools like PhotoDNA and Google's CSAI match tool available for identifying this content. A challenge I hear raised frequently is identifying and removing novel images that have not already been hashed.**
   a. **What would it take to develop better technology to accurately identify and limit the spread of novel CSAM images?**
   b. **Are there interventions from Congress that would facilitate identification of CSAM?**

[6] https://newsroom.tiktok.com/en-gb/enhancing-the-live-community-experience-with-new-features-updates-and-policies

c. Based on your company's experience trying to address online sexual exploitation and abuse of minors, are there areas where Congress could be helpful in tackling this problem?

The success of computer vision models depends on the quality and representativeness of the training data for the models. This is a highly complex issue given the underlying subject matter, and we welcome further dialogue on this extremely important and sensitive topic.

4. **AI models are making it easier to develop synthetic CSAM. These are either altered images of real people, or wholly synthetic individuals. Policymakers are grappling with what this will mean for law enforcement efforts to hold perpetrators accountable and identify children who are being harmed. In addition to processing a higher volume of Cybertips, investigators will have the added challenge of determining whether the victim in the scenario is in fact a real person. And cases are already being reported where AI generative technologies are being employed to facilitate the grooming and sextortion of minor victims.**
   a. **What are you doing to identify and remove AI-generated CSAM on your services?**
   b. **Do you flag for NCMEC if you perceive the CSAM to be AI-generated?**
   c. **How prevalent is this kind of content?**
   d. **How do you anticipate the rise of AI-generated CSAM will impact NCMEC's ability to process and refer Cybertips to law enforcement?**
   e. **Recently, A.I.–generated explicit images of a major pop superstar were distributed widely online without her consent. That story drew attention to a growing problem over the last year facilitated by AI tools: the generation of deepfake, nonconsensual, sexually explicit imagery of everyday people, including our young people. Will you commit to reporting on the prevalence of this new problem and the steps your company is taking to address this horrendous abuse?**
   f. **Are there technical or legal barriers that your company has identified preventing thorough red teaming of AI models to ensure they do not generate CSAM?**

TikTok's approach to combatting AI-generated CSAM begins with our Community Guidelines[7] and the strict policies we have for both Synthetic and Manipulated Media[8] and Youth Exploitation and Abuse[9]. In the context of child safety, we updated our Community Guidelines to make clear that we do not allow any AI-generated content of a real child. Furthermore, we have maintained our zero tolerance policy against CSAM, nudity, grooming, sextortion, solicitation, pedophilia, and physical or psychological abuse of young people, including content that is real, fictional, digitally created, and shown in fine art or objects.

[7] https://www.tiktok.com/community-guidelines/en/
[8] https://www.tiktok.com/community-guidelines/en/integrity-authenticity/?cgversion=2023#3
[9] https://www.tiktok.com/community-guidelines/en/safety-civility/?cgversion=2023#4

Our comprehensive Community Guidelines provide the foundation for our moderation strategies. Our CSAM detection techniques cover our strict policies inclusive of AI-generated content, and our Trust and Safety teams are actively investing in general AI-generated content detection techniques, from which we may further distinguish between permitted and prohibited content. Currently, we flag all content that we confirm to be CSAM and expeditiously report to NCMEC. This includes Perceived First Person CSAM (PFP CSAM) as well as anything possibly computer or AI-generated or altered.

With respect to barriers around thorough red-teaming, CSAM is contraband, and therefore AI-generated CSAM must also be treated as contraband. Red-teaming, or attempting to jailbreak a tool to see if it can generate undesired content, is high risk in the context of CSAM given the potential liability to employees due to the lack of safe harbor laws or affirmative defenses.

5. **How companies choose to allocate their resources illustrates their true priorities.**
    a. **What percentage of your company's budget is dedicated to addressing child safety on your platform?**
    b. **What process or assessment of risk on the platform informed that figure?**
    c. **How many layers of leadership separates your trust and safety leaders from you?**

TikTok currently has more than 40,000 trust and safety professionals working to protect our community. We expect to invest more than two billion dollars in trust and safety efforts in 2024, with a significant part of that investment in our U.S. operations.

TikTok's CEO regularly speaks with its trust and safety leadership and they have an open line of communication. As of February 2024, TikTok's head of global trust and safety reports directly to the CEO, and there is one formal layer of leadership between TikTok's U.S. head of trust and safety and the CEO.

6. **The companies represented at the hearing have the money and resources to hire teams of Trust & Safety professionals and build bespoke tools to aid with content moderation and integrity work as well as the detection of content like CSAM on their services. This is not necessarily the case for the rest of the tech sector. These are industry-wide problems and will demand industry-wide professionalization and work.**
    a. **What is TikTok currently doing to support access to open-source trust & safety tools for the broader tech ecosystem?**
    b. **And if TikTok is not doing anything now, will you commit to supporting the development of these kinds of resources?**

TikTok agrees these are industry-wide issues and supports the development of additional tools that can be used by companies at all stages of their development. We financially support organizations like the Tech Coalition and others that work to create accessible resources and to aid in the development and access to detection systems and remain committed to partnering with others across the broader tech ecosystem to develop additional resources.

7. **One necessary element of keeping our kids safe is preventing harms in the first place. The National Center for Missing and Exploited Children partnered with the White House, the Department of Justice, and the Department of Homeland Security to create "The Safety Pledge" initiative to combat online child exploitation in September 2020. I understand more government backed public awareness campaigns are being developed.**
    a. **Are you partnering with the federal government to distribute health and safety resources to young people?**
    b. **What are you proactively doing to educate the minors that use your services about online health and safety?**

We are committed to creating a safe platform for our community, and believe this work is especially important for our teenage users. People should be able to come to TikTok to express themselves creatively and be entertained in a safe and inclusive environment. Our comprehensive approach to teen safety includes robust policies, innovative technologies, in-app features, and educational resources. We also work collaboratively with industry partners, non-profits, academics, and governments, to identify, implement, and share innovative solutions to better protect teens online.

On our Safety Center, we offer a number of tools and controls to help teens manage their experience. We provide guides, including our well-being guide[10], to share more about our approach to safety, privacy, and security on TikTok. We also offer helpful information for parents, caregivers, and new users. Additionally, listening to the experience of teens is one of the most important steps we can take to build a safe platform for teens and their families. We launched TikTok's global Youth Council, where we will listen to the experiences of those who directly use our platform and be better positioned to make changes to create the safest possible experience for our community.

TikTok is continuously working to provide a safe app experience for our community, and we aim to be a leader in this area. We recognize, however, that technology is ever-evolving and that we need to be prepared to address unexpected trends and challenges as they arise. Beyond our efforts with NCMEC, TikTok works with a variety of global partners on minor safety efforts:

- We are an active member of the Tech Coalition, a global alliance of technology companies to protect children online. In addition to being on its board, we also co-led the multi-stakeholder forum on Minor Financial Sextortion held last year, and we chair committees within the organization to advance the fight against online child sexual exploitation and abuse.
- We are part of the WePROTECT Global Alliance, the largest and most diverse multi-sector alliance dedicated to ending online child sexual exploitation.
- Internet Watch Foundation (IWF) is a vital partner for TikTok in our work to counter online child sexual abuse and exploitation. TikTok accesses IWF's URL and keyword list and, via NCMEC, its hash database of known child sexual abuse images. In addition to its

[10] https://www.tiktok.com/safety/en/well-being-guide/

frontline work, IWF provides insight on new and emerging trends and acts as a convener for key stakeholders.

- We worked with ConnectSafely—a nonprofit dedicated to educating users of connected technology about safety, privacy, and security—to develop a TikTok-specific guide for parents and teens.
- TikTok's Top 10 Tips for Families guide for the Family Online Safety Institute offers information on several tools to help teens manage how they interact with other users and who can see their videos. It includes information about privacy restrictions, content, comments, and messages.

Additionally, TikTok is part of the alliance in support of the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse. This alliance brings together governments across the world, including the United States; digital platforms and other private companies; and civil society and intergovernmental organizations. The objective is to collaborate across sectors and industries to address online threats to minors. TikTok also works with a group of non-governmental organizations as part of a CSAM intervention project designed in part to help direct users to appropriate resources.

Beyond the partnerships with leading organizations in the space, we see others finding great success in leveraging TikTok as a platform to spread critical safety messages. For example, Thorn launched an in-app campaign called NoFltr, which facilitated a conversation between youth and adults on the sharing of nude images and consent. We also worked in collaboration with Thorn researchers and our policy team to inform our policy and features related to their findings.

We engage in these partnerships because online threats are complex and dynamic. No single company or government can solve these problems in a vacuum. We look forward to continuing our own work, as well as our collaborations, to make a meaningful difference in keeping our community safe.

8. **Sextortion has become increasingly prevalent. Offenders may use grooming techniques or basic trickery to manipulate victims into providing nude or partially nude images of themselves, which are then used to coerce victims into sending more graphic images and videos or pay a ransom. These criminals often threaten to post the images or sensitive images publicly or send them to the victim's friends and family if the child does not comply. From May 2022 to October 2022, U.S. law enforcement and NCMEC witnessed an alarming increase in CyberTips and reports where minors have been sextorted for money. Many young boys, including in California, have committed suicide out of desperation, leaving their loved ones devastated.**
    a. **How is your company responding to the growing threat of financial sextortion?**
    b. **What methods are in place to detect and disrupt this type of abuse in real time?**
    c. **What kind of user education and awareness are you engaged in?**

d. Are you aware of a higher prevalence of sexual extortion or abuse against certain demographics among young users? If not, will you commit to studying this issue and making that kind of information available to improve public education and protection measures?

TikTok's Community Guidelines prohibit sexual exploitation or gender-based violence, including non-consensual sexual acts, image-based sexual abuse, sextortion, physical abuse, and sexual harassment. Although not required by law, TikTok reports sextortion-related content to NCMEC for prioritized NCMEC review and forwarding to law enforcement. TikTok also participates in NCMEC's Take It Down[11] service, which is a free service that can help process requests to remove or stop the online sharing of nude, partially nude, or sexually explicit images or videos taken when someone is under 18 years old.

TikTok has restrictions in place to help disrupt and minimize this sort of content. For example, users under 16 do not have access to direct messaging. Additionally, if the content shared in direct messaging is detected by our CSAM related models, the content will be blocked and not will not be delivered.

TikTok is a member of the Tech Coalition[12], which helps facilitate high-impact information, expertise, and knowledge sharing across industry to disrupt and help prevent online CSEA, including creating and expanding robust systems and processes for information and threat sharing related to exploitative or predatory behaviors. In turn, the Tech Coalition financially supports academic research on these important issues through its Safe Online Research Fund[13]. In addition to being on the board of the Tech Coalition, we also co-led the multi-stakeholder forum on Minor Financial Sextortion held last year, and we chair committees within the organization to advance the fight against online child sexual exploitation and abuse.

9. Young people need to be at the center of regulatory discussions, and they need to be at the table as products and services they use are designed.
   a. Are you engaging young adults and youth in your conversations and policies around Trust and Safety on the platform?
   b. How do you proactively keep up to speed with the most pressing issues facing young people online?

At TikTok, we prioritize the safety and well-being of our community. As we work to provide a safe and inclusive place for everyone, we actively seek input and advice from individual experts and nonprofit organizations. Through ongoing engagement, we work to ensure our policies and processes are informed by a diversity of perspectives, expertise, and lived experiences. By bringing together different voices, we aim to create a safe platform for everyone, especially those who may be more vulnerable to online harm.

---

[11] https://takeitdown.ncmec.org/
[12] https://www.technologycoalition.org/what-we-do
[13] https://www.technologycoalition.org/newsroom/tech-coalition-safe-online-research-fund-announces-additional-funding-of-us-500k-to-select-existing-grantees-for-research-extension-product-development-innovation

To do this, we consult with a number of NGOs, academics, and civil society members to better inform our work. In addition, we host global Advisory Councils[14], and have formally established a Youth Council[15] that directly engages with teens to solicit their input.

10. **For many children, an open dialogue about their internet habits is a best practice, and healthy. But not every child has a parent or a caregiver that is looking out for their best interest. For many kids who are abused, a caregiver or parent is their abuser. Additionally, for many young people, their parents' knowledge of their sexual orientation or their interest in exploring it, fundamentally puts them in jeopardy. Solving for these different needs across our young people at the scale of social media and internet applications is really vital.**
    a. **How have you designed your parental tools with this dynamic in mind?**

We are committed to creating a safe platform for our community, and believe this work is especially important for our teenage members. People should be able to come to TikTok to express themselves creatively and be entertained in a safe and inclusive environment. Our comprehensive approach to teen safety includes robust policies, innovative technologies, in-app features, and educational resources. We also work collaboratively with industry partners, non-profits, academics, and governments, to identify, implement, and share innovative solutions to better protect teens online.

Listening to the experience of teens is one of the most important steps we can take to build a safe platform for teens and their families. It helps us avoid designing teen safety solutions that may be ineffective or inadequate for the actual community they're meant to protect, and it brings us closer to being a strong partner to caregivers as we can better represent teens' safety and well-being needs. We launched TikTok's global Youth Council[16], where we will listen to the experiences of those who directly use our platform and be better positioned to make changes to create the safest possible experience for our community. In a similar way to how we engage regularly with more than 50 academics and leading experts from around the world through our Content and Safety Advisory Councils[17], the new Youth Council will provide a more structured and regular opportunity for youth to provide their views.

We have launched several initiatives aimed at supporting teens' digital journeys and helping ensure that online experiences play a positive role in how younger users express themselves, discover ideas, and connect. In addition to offering a range of safety and privacy controls that empower users to decide who they share content with, TikTok provides even stronger proactive protections to safeguard our teen users, and we have consistently introduced changes to support age-appropriate experiences on our platform. For instance, accounts registered to teens under 16 are set to private by default, and their content is ineligible for recommendation to people they do not know.

[14] https://www.tiktok.com/transparency/en/advisory-councils/
[15] https://newsroom.tiktok.com/en-us/updating-family-pairing-and-establishing-tiktoks-youth-council
[16] https://newsroom.tiktok.com/en-us/updating-family-pairing-and-establishing-tiktoks-youth-council
[17] https://www.tiktok.com/transparency/en/advisory-councils/

TikTok's Family Pairing features let a parent or guardian link their TikTok account to their teens to enable a variety of content and privacy settings. Within Family Pairing, we provide tips for caregivers[18] that we developed in collaboration with teens. These outline the support teens would like and their suggestions on how to approach conversations about digital literacy and safety. We encourage caregivers to discuss the Family Pairing features with their teens and explain why they choose to turn them on. Even without Family Pairing enabled, parents can help their teens enable TikTok's Screen Time offerings, including Daily Screen Time and Restricted Mode, which are protected by a passcode set by the parent or guardian.

Last year we also heard from parents and caregivers that they'd like more ways to customize the topics their teen may prefer not to stumble upon, as every teen is unique and caregivers are often closest to their teen's individual needs. To adapt this feature for Family Pairing, we engaged with experts, including the Family Online Safety Institute[19], on how to strike a balance between enabling families to choose the best experience for their needs while also ensuring we respect young people's rights to participate in the online world. Therefore, by default, teens can view the keywords their caregiver has added and we believe this transparency can also help to prompt conversations about online boundaries and safety. The keywords caregivers add will be a personalized layer on top of our Content Levels[20] system, which already helps to keep content with more mature or complex themes from reaching audiences between ages 13-17.

[18] https://newsroom.tiktok.com/en-us/new-family-pairing-resources-offer-digital-safety-advice-from-teens
[19] https://www.fosi.org/
[20] https://newsroom.tiktok.com/en-us/more-ways-for-our-community-to-enjoy-what-they-love

TikTok's Responses to Questions for the Record from Senator Welch
Submitted March 7, 2024

1. **Please explain TikTok's decision to not offer full end-to-end encryption for direct messaging.**

Direct messages on TikTok are encrypted at rest and while in transit. End-to-end encryption is not currently available. We place a premium on ensuring that our younger users have a safe experience by default on TikTok. Like many companies, we maintain the ability to decrypt user data in response to valid legal process[1] and to enforce our Community Guidelines[2], and we regularly publish Transparency Reports[3] to provide visibility into this work. We maintain internal controls to ensure that only personnel with proper authorization and a demonstrated need to perform their job have access to certain decrypted data like contact information or direct messages.

2. **Since TikTok does not offer end-to-end encryption for direct messages, how can users be sure that foreign governments or bad actors are not snooping on their private conversations?**

TikTok does not believe the use of encryption increases TikTok's cybersecurity vulnerabilities. TikTok uses industry-standard encryption to protect sensitive user data. We encrypt sensitive user data in transit and at rest. Data can only be decrypted with a key that is generated and managed by our key management service, which is operated by TikTok's established subsidiary, US Data Security Inc. ("USDS").

3. **How much of your content moderation is managed by artificial intelligence?**

TikTok removes violative content proactively using both automated and manual processes, as well as removing violative content reported by users. Videos uploaded to TikTok are initially reviewed by TikTok's automated moderation technology, which aims to identify content that violates the Community Guidelines before it is distributed across the TikTok platform and displayed to users.

We proactively remove content that violates our Community Guidelines. For example, in 2023 Q3, we removed more than 136 million violative videos globally, which accounts for just under 1% of total published videos over that period. The vast majority (96%) were removed proactively, before they were reported to us. Approximately 65% of our removals were by automation.

[1] https://www.tiktok.com/legal/page/global/law-enforcement/en
[2] https://www.tiktok.com/community-guidelines/en/
[3] https://www.tiktok.com/transparency/en-us/reports/

4. **Is it your view that artificial intelligence can replace human judgment in identifying and removing false or harmful content? If not, when is human judgment necessary?**

TikTok's automated technology systems look at a variety of signals across content, including keywords, images, titles, descriptions, and audio, and continuously learn and adapt based on the data in each video and the moderation decisions that TikTok's human moderators ultimately make based on TikTok's Community Guidelines and related policies. If TikTok's machine-based filters identify a potential violation, the automated moderation system will either pass it on to TikTok's safety teams for further review or remove it automatically. Automatic removal is applied if there is a high degree of confidence that the content violates the Community Guidelines and where violations are most clear-cut. Automatic removals are subject to the user's ability to appeal that determination.

When TikTok's automated moderation systems identify potentially problematic content but cannot make an automated decision to remove it, they send the content to TikTok's safety teams for further review. To support this work, TikTok has developed technology that can identify risky or suspicious items — for example, weapons — in video frames, so that content moderators can carefully review the video and the context in which it appears. This technology improves the efficiency of TikTok's moderators by helping them more adeptly identify violative images or objects, quickly recognize violations, and make decisions accordingly.

In keeping with its commitment to ensuring fairness, TikTok notifies community members if and why their content was removed. The community member can then appeal the decision if he or she believes their content was erroneously removed, and TikTok allows community members to submit specific feedback on why they disagree with the decision to remove the content.

5. **How have your Trust & Safety teams been trained on how to handle false or illegal AI-generated content?**

We identify and remove any content that violates our Community Guidelines, regardless of whether it was created or altered by AI. This is addressed through a combination of proactive technology, user reports, and flags from trusted partners.

AI-generated content ("AIGC") brings new challenges around misinformation in our industry, which we've proactively addressed with firm rules and new technologies. We don't allow manipulated content that could be misleading, and we also require creators to label any realistic AIGC and launched a first-of-its-kind tool to help people do this.

6. **How does TikTok plan on addressing the large amount of disinformation that could be spread on its platform during the 2024 election?**

Thousands of trust and safety professionals work alongside technology to enforce our Community Guidelines. We are committed to consistently enforcing our rules to fight misinformation, covert influence operations, and other content and behavior that platforms see more of during elections.

- **Countering misinformation**: We invest in media literacy as a counter-misinformation strategy as well as technology and people to fight misinformation at scale. This includes specialized misinformation moderators with enhanced tools and training, and teams on the ground who partner with experts to prioritize local context and nuance. We partner with 17 global fact-checking organizations, who assess the accuracy of content in over 50 languages so that our moderators can apply our misinformation policies accordingly. We added three new global fact-checking partners in 2023, and will continue to expand our fact-checking program this year.
- **Deterring covert influence operations:** We know that deceptive actors try to target online platforms during elections, and we remain vigilant[4] against covert influence operations. We have dedicated experts working to detect, disrupt, and stay ahead of deceptive behaviors. We report the removals of covert influence networks in our quarterly Community Guidelines Enforcement Reports. In the coming months, we'll introduce dedicated covert influence operations reports to further increase transparency, accountability, and sharing with the industry. We provide information about how we assess this behavior at our Transparency Center[5].
- **Tackling misleading AI-generated content:** AI-generated content ("AIGC") brings new challenges around misinformation in our industry, which we've proactively addressed with firm rules and new technologies. We don't allow manipulated content that could be misleading, including AIGC of public figures if it depicts them endorsing a political view. We also require creators to label any realistic AIGC and launched a first-of-its-kind tool to help people do this. As the technology evolves in 2024, we'll continue to improve our policies and detection while partnering with experts on media literacy content that helps our community navigate AI responsibly.

[4] https://www.tiktok.com/transparency/en-us/countering-influence-operations/
[5] https://www.tiktok.com/transparency/en-us/

TikTok's Responses to Questions for the Record from Senator Butler
Submitted March 7, 2024

1. **Family and parental control tools: I was glad to hear that you have spent time talking with parents and what their families need from your products. I was also glad to hear your companies have a Family Center, or other similar tools, to give parents more insight and control over how their children are using your platforms and apps.**
    a. **How do you advertise this feature to parents?**
    b. **Can you share data on how many Family Center/parental tools users there are in proportion to total minors on your platforms and products?**

TikTok has invested significant resources in promoting Family Pairing, including utilizing both earned and paid media such as advertisements, billboards, and local news. Family pairing has also been promoted to users via in-app notifications and on our Guardian's Guide[1].

As a privately held company, TikTok does not disclose detailed statistics about its users or adoption of specific features.

[1] https://www.tiktok.com/safety/en/guardians-guide/

TikTok's Responses to Questions for the Record from Senator Grassley
Submitted March 7, 2024

1. **Current law requires that a provider of a report of suspected CSAM to the National Center for Missing and Exploited Children's (NCMEC) CyberTipline preserve "any visual depictions, data, or other digital files that are reasonably accessible and may provide context or additional information about the reported material or person" for a minimum of 90 days. 18 U.S.C. 2258A(h)(1-2). The recent explosion of suspected abuse has presented unprecedented challenges for law enforcement to follow up on leads before companies discard or delete essential data and information. There is nothing preventing tech companies from preserving relevant material beyond the statutorily-mandated 90-day period.**

    a. **How long does TikTok voluntarily preserve and retain data contained in and related to its reports to the CyberTipline?**

TikTok proactively preserves and retains data related to the user and incident reported in a NCMEC report for 180 days.

    b. **The massive influx of reports to the CyberTipline naturally results in law enforcement entities having to conduct and finish investigations beyond 90 days of an initial report to the CyberTipline. Retaining relevant information for longer periods could significantly advance law enforcement's ability to thoroughly investigate leads. If TikTok only preserves and retains this information for the minimum 90-day period, why does it do so when preserving this data longer could significantly enhance and prolong law enforcement's ability to investigate and prosecute child predators?**

Not applicable - see response to Question 1(a).

    c. **Please confirm if TikTok stores and retains the following information relating to reports to the CyberTipline:**
        i. **IP addresses**
        ii. **Screen Names**
        iii. **User Profiles**
        iv. **Associated Screennames (by IP address and associated emails)**
        v. **Email addresses**
        vi. **Geolocation data**

TikTok stores and retains the following information related to its reports to the CyberTipline, to the extent we have such information for the user: IP address, username, subscriber information, associated username, email address, and registration location.

d. If TikTok does not retain or store any of the above types of information in question (c), please explain why.

Not applicable - see response to Question 1(c).

e. Please list any other information TikTok retains and preserves for law enforcement purposes not listed above in question (c).

TikTok also retains and preserves all reported user content.

f. Does TikTok flag screennames and associated email addresses to suspected accounts that violate TikTok's terms of service?

Any account that violates our policies and is reported to NCMEC will be removed along with associated accounts. We also block their device from creating new accounts.

2. How does TikTok prioritize urgent requests for information from law enforcement and what is TikTok's response time to urgent requests?

TikTok's Law Enforcement Response Team ("LERT") prioritizes requests marked as urgent by aiming to respond within 1 business day. Otherwise, LERT follows internal prioritization policies based on the nature of the crime described in law enforcement requests. The nature of case categories that are prioritized include same day responses for human trafficking and within three business days for cases involving child exploitation.

TikTok's Emergency Response Team processes emergency disclosure requests ("EDRs") from law enforcement with a team of specialists available at all times. The team assigns the emergency request within 15 minutes and aims to respond to law enforcement within 1 hour.

3. What is TikTok's average response time to service of legal process from law enforcement for CSAM-related information?

On average in 2023, TikTok responded to legal process requests from law enforcement related to child exploitation investigations within approximately 15 days, dependent on the type of legal process and amount of data required by disclosure.

4. In 2023, the tech industry as a whole slashed more than 260,000 jobs. And in the first four weeks of this year, another 25,000 jobs were cut.
   a. For each year, between 2018 and 2023, how many U.S. based employees did you have at TikTok?
      i. Of these employees, how many were sponsored on H-1B visas?
      ii. For each year, between 2018 and 2023, how many H1-B visa applications did TikTok submit?
   b. For each year, between 2018 and 2023, how many employees based outside the U.S. did you have at TikTok?

i. **Of these employees, how many were based in China?**

c. **For each year, between 2018 and 2023, how many employees in total did TikTok terminate, fire, or lay off?**
   i. **Of these employees, how many were based in the United States?**
   ii. **Did TikTok fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?**
   iii. **Were any duties and/or functions previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?**

d. **For each year, between 2018 and 2023, how many employees performing work related to child safety did TikTok terminate, fire, or lay off?**
   i. **Of these employees, how many were based in the United States?**
   ii. **Did TikTok fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?**
   iii. **Were any duties and/or functions (specifically relating to child safety) previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?**
   iv. **How have layoffs impacted TikTok's ability to protect children on its platforms?**
   v. **Does TikTok have any plans to increase staff responsible for child safety operations or otherwise optimize its child safety operations?**

TikTok is working diligently on a response to this question and its subparts. Review of the associated data is not complete at the date of submission of these responses, so TikTok will follow up with Senator Grassley to share relevant data.

5. **On January 30, 2024, the Tech Transparency Project (TTP) published an article on their website called, "Meta Approves Harmful Teen Ads with Images from its Own AI Tool". In summary, TTP, using Meta's "Imagine with Meta AI" tool generated inappropriate images such as young people at a pill party or other vaping. These images with text were submitted to Facebook as advertisements targeting users between ages 13-17 in the United States. TTP reported that Facebook approved the advertisement, despite it violating its own policies, in less than five minutes to run on the following platforms: Facebook, Instagram, Messenger, and Meta Quest. Meta. Over the course of a week, TTP submitted the advertisements with the same end result: Facebook approving them. TTP reported that they canceled these advertisements before their scheduled publication, but it illustrated the repeated failures of Facebook to properly moderate content. This is just one example of what other non-government organizations and others have uncovered across social media platforms.**

a. How often a month do TikTok employees conduct quality checks on TikTok's policies and safeguards for child accounts?

We work hard every day to provide a safe, trustworthy and vibrant experience for users and maintain a set of Community Guidelines that includes rules and standards for using TikTok, including around youth safety and well-being. The guidelines apply to everyone and everything on our platform. They are informed by international legal frameworks and industry best practices. TikTok also regularly consults our community, safety and public health experts, and our regional Advisory Councils to inform our policy development, stay on top of evolving issues, and address emerging risks and potential harms that may occur from new behaviors.

b. In which departments, components, or units of the company does TikTok have staff dedicated to performing this type of work?

TikTok's community is protected by over 40,000 global trust and safety professionals who help maintain and enforce our robust Community Guidelines[1], Terms of Service[2] and Advertising Policies[3], which apply to all content on our platform.

c. How many employees make up these departments, components, or units?

As noted in our response to Question 5(b), TikTok currently has more than 40,000 trust and safety professionals working to protect our community.

d. If a violation is found, what action is taken, and how quickly is action taken?

We aim to remove content or accounts that violate our guidelines before they are viewed or shared by other people in order to reduce potential harm. In our latest transparency report[4] covering Q3 2023, the removal rate for violative content was 96.1% before a single view and 90.6% within 24 hours of posting.

Content on TikTok first goes through technology that reviews it against our Community Guidelines. If content is identified as a potential violation, it will be automatically removed, or flagged for additional review by our trust & safety team. Additional review will occur if a video gains popularity or has been reported. Community members can report[5] violations in-app and on our website. Our quarterly Community Guidelines Enforcement Reports[6] provide insight into our enforcement efforts, showing how we continue to uphold trust, authenticity, and accountability.

---

[1] https://www.tiktok.com/community-guidelines/en/
[2] https://www.tiktok.com/legal/page/us/terms-of-service/en
[3] https://ads.tiktok.com/help/article/tiktok-advertising-policies-ad-creatives-landing-page-prohibited-content?lang=en
[4] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/
[5] https://support.tiktok.com/en/safety-hc/report-a-problem
[6] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/

6. **Social media companies claim they are investing in company components dedicated to safety, and that their platforms are safe for children. However, children continue to be exploited daily across these platforms.**
    a. **What have TikTok's revenue and profit figures been for the last three years (2021-2023)? Please provide figures broken out per year. Do not provide percentages.**
    b. **How much has TikTok spent in advertising for the last three years (2021-2023), broken out per year?**
    c. **How much of TikTok's resources spent on advertising has been devoted to advertising TikTok's safety initiatives and efforts for the last three years (2021-2023), broken out per year?**
    d. **To get an understanding of how your company has invested and plans to invest in its components dedicated to child safety functions, what are the annual budgets for TikTok's child safety-related components for the last three years (2021-2023)?**
    e. **What is the current anticipated (2024) budget for TikTok's child safety-related components?**
    f. **Provide the number of staff employed in TikTok's child safety-related components for the last three years (2021-2023).**
    g. **How much is that compared to TikTok's other components for the same period? (Please provide a breakout per year. Do not provide percentages.)**
    h. **How many staff are currently employed in TikTok's child safety-related components?**
    i. **What are the roles, responsibilities, and functions of TikTok's child safety-related components?**
    j. **Are any other components responsible for the monitoring of CSAM on TikTok's platform(s)?**
    k. **What, if any, third parties does TikTok employ or contract with to address CSAM material on its platforms?**
        i. **What are the roles and responsibilities of these third parties?**
        ii. **What is the breakdown of cost per third party over the last three years (2021-2023)?**

As a privately held company, TikTok does not disclose revenue or detailed budget breakdowns. We expect to invest more than two billion dollars in trust and safety efforts in 2024, with a significant part of that investment in our U.S. operations.

TikTok's U.S. community is protected by over 40,000 global trust and safety professionals. Within this, TikTok's content moderators work alongside our automated moderation systems to take into account additional context and nuance which may not always be picked up by technology. Our moderators are trained across all Community Guidelines violations, including youth safety issues. Some violations require further work by specialized moderators in specialized queues. We are continuing to hire for our Trust and Safety team globally, with 1,000+ open roles.

TikTok has invested significant resources in promoting safety features such as Family Pairing, including utilizing both earned and paid media such as advertisements, billboards, and local news. Family pairing has also been promoted to users via in-app notifications and on our Guardian's Guide[7].

7. **Of all reports sent by TikTok to the National Center for Missing and Exploited Children, how many reports were self-generated from victim users for the last three years (2021-2023)? Please provide the actual number of self-generated reports in addition to the total number of reports (including those that were not self-generated). In addition, please provide a break-down of the self-reporters by age.**

We currently do not track if an incident reported to NCMEC was solely initiated by the victim-user.

8. **What is TikTok's policy or protocol with respect to law enforcement accessing user data and subsequent notification to users of law enforcement accessing their data?**
    a. **Do certain crimes such as drug trafficking or child exploitation affect TikTok's decision to notify a user whose data is accessed by law enforcement?**
    b. **Do certain requests such as a subpoena or search warrant affect TikTok's notification protocol? If so, what are they?**
    c. **If TikTok does notify users of law enforcement accessing their data, why does TikTok find this necessary?**

TikTok does not currently notify U.S. users of law enforcement requests.

9. **The National Center for Missing and Exploited Children has indicated that reports from social media companies tend to lack actionable information causing law enforcement to be burdened with incomplete information. How comprehensive are TikTok's reports to NCMEC? What challenges is TikTok experiencing on the collection of user data and other information to include in its reports to NCMEC? What actions is TikTok taking to make its reports more comprehensive and useful to law enforcement?**

In 2022, NCMEC informed TikTok that 83% of its reports were 'actionable' by law enforcement. We provide the same information in all of our reports regarding the offender and the content. In 2022, we received only one request from NCMEC[8] to take down content, and removed it in a matter of hours.

[7] https://www.tiktok.com/safety/en/guardians-guide/
[8] https://www.missingkids.org/content/dam/missingkids/pdfs/2022-notifications-by-ncmec-per-esp.pdf

10. Has TikTok applied to join Project Lantern to work with industry peers tackling online child predators and the reporting of Child Sexual Abuse Material to law enforcement authorities?

a. If not, why hasn't TikTok applied for participation in Project Lantern when doing so could improve law enforcement's ability to effectively investigate these crimes?

TikTok has applied to join Lantern and has been tentatively approved for membership.

11. Mr. Chew, you testified, "we have not been asked for any data by the Chinese government and we have never provided it." That same day, Director Wray testified about TikTok before the House Select Committee on the Chinese Communist Party. In describing his concerns about TikTok, he stated, "the most important starting point is the role of the Chinese government. The app's parent company is effectively beholden to the Chinese government and that is what in turn creates a series of national security concerns in the PRC government's ability to leverage that access or that authority."

a. Has any member of the Chinese government or Chinese Communist Party asked ByteDance or any of ByteDance's other subsidiaries for access to U.S. company, U.S. person, or U.S. government data? If so, what data has the Chinese government or Chinese Communist Party requested?

TikTok does not offer the TikTok app for download in mainland China. TikTok Inc. has not been asked by the Chinese government for U.S. user data. TikTok discloses on a regular basis in its Information Requests Reports[9] the volume and type of requests for user information received from governments and law enforcement agencies, and whether the data was disclosed or presented.

b. What information or data related to the United States companies, United States government, or United States users does TikTok share with ByteDance?

TikTok's data sharing practices are described in TikTok's U.S. Privacy Policy[10] and U.S. Children's Privacy Policy[11]. Subject to the further limitations on protected U.S. user data that are described below, access by TikTok-affiliated corporate group entities is governed by the principles set out in ByteDance's data access policies.

In addition to the safeguards imposed by the data access policies, TikTok has been engaged in ongoing confidential discussions with CFIUS regarding measures that will significantly limit TikTok's affiliates' access to protected U.S. user data. In connection with these discussions, TikTok formed a new subsidiary, TikTok U.S. Data Security Inc. ("USDS"), which will be

[9] https://www.tiktok.com/transparency/en-us/information-requests-2023-1/
[10] https://www.tiktok.com/legal/page/us/privacy-policy/en
[11] https://www.tiktok.com/legal/page/global/childrens-privacy-policy/en

tasked with managing all business functions that require access to protected U.S. user data and the deployment and functioning of the TikTok app and TikTok platform in the United States.

c. What information or data related to United States companies, United States government, or United States users does TikTok share with ByteDance's other subsidiary companies?

TikTok's parent company, ByteDance Ltd., has China-based subsidiaries including Beijing Douyin Information Services Co., Ltd. The 1% stake in that entity does not give the Chinese government any right to influence the operations of TikTok. Please see the corporate structure diagram at ByteDance.com, which illustrates the structural separation between Beijing Douyin Information Services, Co. Ltd. and all TikTok entities.

TikTok has imposed data access policies to help ensure that adequate safeguards are in place to protect personal information. If a person has no business need to access protected U.S. user data, they are not afforded such access under the terms of those policies. In addition to these safeguards, USDS is undertaking efforts that are unprecedented among our peer group to build a secure environment for protected U.S. user data.

12. On February 4, CBS News reported that videos on TikTok were providing migrants (including large numbers of Chinese migrants) with instructions on how to hire a human smuggler and enter the United States illegally through a gap in the fence along the California border.

a. Is this a violation of TikTok's policy?

Yes, this type of content would violate TikTok's Community Guidelines. TikTok prohibits content that seeks to promote or facilitate criminal activities, including human smuggling. We do not allow:

- facilitating or promoting human trafficking and human smuggling activities
- requesting support for being smuggled illegally into a different country
- instructional content on how to illegally cross the border

b. Why hasn't TikTok already removed this material from its platform? Will TikTok commit to removing this material from its platform?

As mentioned, this type of content would violate our Community Guidelines. When we identify content that violates our policies, it is removed from the platform. Our policies relating to human exploitation allow certain content, such as content that expressing a desire to migrate to another country, or showing a migrant's journey (as long as it does not explicitly show the involvement of smugglers in their journey or provide instructions on how to illegally cross the border). We report this type of content to law enforcement in certain circumstances, and we also work with

third-party intelligence firms to bolster our defenses against the efforts of bad actors on the platform and make reports to law enforcement as appropriate.

c. How can TikTok reconcile its stated commitment to protect children with allowing its platform to facilitate a crisis which is especially dangerous to children and could lead to their trafficking, abuse, and exploitation?

We are committed to upholding human rights and preventing the platform from being misused to enable any exploitative activities. We do not allow human exploitation on our platform, including trafficking, smuggling, forced labor, or underage marriage. We do understand how important it is for survivors who have experienced exploitation to share their stories and for migrants and refugees to be able to describe challenges they faced, so we seek to create a supportive space to do so (provided it does not violate any of our Community Guidelines, including those noted in our response to Question 12(a)).

We recognize our responsibility to help maintain a safe and welcoming environment for our community. Our Law Enforcement Response Team discloses relevant user data in response to valid legal requests from law enforcement agencies, including the U.S. Department of Homeland Security. TikTok may also share content or account information directly with law enforcement in the absence of a request when it believes in good faith that there is an emergency involving imminent harm or risk of death or serious physical injury to a person. Additionally, we have a robust law enforcement outreach team that is dedicated to meaningful engagement with law enforcement officers across the federal, state, and local levels. Since 2021, this team has trained over 13,000 U.S. law enforcement officers on our platform and company. We maintain an open dialogue for questions and answers through this program, and provide resources and contacts for questions and emergencies.

TikTok's Responses to Questions for the Record from Senator Lee
Submitted March 7, 2024

1. **The 2022 Thorn Report identified TikTok as the #3 platform where minors reported having an online sexual encounter (12 percent), and with 8 percent of minor TikTok users having an online sexual encounter with someone they believed to be an adult. Amongst minors who share self-generated CSAM, 52 percent say they do so with people they only know from online interactions. What is TikTok doing to cease being one of the most preferred platforms for predators to sexually exploit children?**

We are deeply committed to ensuring that TikTok offers a safe and positive experience for people under the age of 18. Minors must be 13 years and older to have an account, with additional age limitations based on local law in some regions. In the United States, there is a separate under 13 TikTok experience[1], which does not permit sharing of personal information and puts extensive limitations on content and user interaction to help keep younger users safe while enjoying TikTok. In this ecosystem, account holders cannot do things like share their videos, comment on others' videos, message with users, or maintain a profile or followers.

Our goal is to provide young people with an experience that is developmentally appropriate and helps to ensure a safe space for self-exploration. We take several steps[2], such as using restrictive default privacy (e.g., setting accounts for 13-15 year olds to private by default) and limiting access to certain product features[3] (e.g., disabling direct messaging and limiting comments for 13-15 year olds, preventing accounts under 18 from being suggested as a connection to others on TikTok, preventing accounts under 18 from going LIVE).

Youth safety is our priority. We do not allow content that may put young people at risk of exploitation, or psychological, physical, or developmental harm. This includes child sexual abuse material ("CSAM"), youth abuse, and exposure to overtly mature themes. If we become aware of youth exploitation on our platform, we will ban the account, as well as any other accounts belonging to the person. If we suspect predatory behavior from a user, their account will be flagged and minor users' videos will be filtered out from their For You page. We also machine moderate DMs for egregious violations (e.g., minor sexual solicitation) to block violative messages from being sent, with the model being stricter for conversations involving a user under 18.

2. **A 2023 survey by Parents Together found that 54 percent of all TikTok users have been exposed to sexual content on TikTok. TikTok restricts certain content from accounts that belong to minors, including removing the direct-messaging function for users 16-years-old and younger. However, the only age verification measure that**

[1] https://newsroom.tiktok.com/en-us/tiktok-for-younger-users
[2] https://support.tiktok.com/en/account-and-privacy/account-privacy-settings/privacy-and-safety-settings-for-users-under-age-18
[3] https://support.tiktok.com/en/account-and-privacy/account-privacy-settings/privacy-and-safety-settings-for-users-under-age-18

TikTok undertakes to ascertain the age of its users is asking new users to enter their birthdate when they open an account. How do you prevent minors from lying about their age when creating an account and accessing the sexual content on your app?

TikTok disagrees with many of the assertions in this question.

TikTok's goal of providing an age-appropriate experience to its users begins with an industry-standard neutral age gate that is consistent with the Federal Trade Commission's ("FTC") guidance for age verification under the Children's Online Privacy Protection Act ("COPPA"). If an individual selects a birthdate that indicates that they are under the age of 13 when creating a TikTok account in the U.S., they are directed to TikTok's under 13 experience, where they can watch a curated library of age-appropriate videos. In addition to being restricted to only certain approved content, users in the under 13 experience cannot access many of the features and functions that are available to users on the 13+ experience. For example, they are not able to post videos on the platform, comment on videos, message other users, maintain a profile or followers, receive ads, or be directed off the TikTok platform.

Beyond age gate, TikTok uses technologies and human moderators, as well as user and third party reporting, to detect and remove users in the 13+ experience who are suspected to be under 13. An account in the 13+ experience that is flagged as being potentially under 13 is routed to a dedicated team of trained moderators who would review the account to determine if it should be banned for not meeting the minimum age requirement. If the moderator makes a determination that the account belongs to a suspected underage user, the account would be removed from the 13+ experience.

Our age assurance measures include a range of measures that are applied on an ongoing basis, rather one time only. In our most recent quarterly report for Q3 2023, we disclosed that we removed 20,864,857 accounts for being suspected of belonging to persons under 13 years of age. Precise numbers are provided on a quarterly basis in our Community Guidelines Enforcement Reports[4] that we make available publicly on our website.

3. **What does TikTok do when you identify a minor who lied about their age when creating an account?**

Accounts that a moderator suspects to be under 13 years of age in the 13+ experience are removed from the platform. In our most recent quarterly report for Q3 2023, we disclosed that we removed 20,864,857 accounts for being suspected of belonging to persons under 13 years of age. Precise numbers are provided on a quarterly basis in our Community Guidelines Enforcement Reports[5] that we make available publicly on our website.

[4] https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement/
[5] https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement/

4. **Predators often contact minors on one platform and then attempt to get the minor to move the conversation to another platform. What does TikTok do to identify these situations, and are these types of interactions reported to law enforcement?**

When a violative conversation or DM is reported, human moderators will review the conversation and escalate to our Child Safety Team. We machine moderate conversations including for grooming, though depending on the nature of the conversation, it may not trigger an intervention since moving off-platform alone is not a policy violation. However, other parts of the conversation may trigger machine intervention.

If applicable, violative conversations will be reported to the National Center for Missing & Exploited Children ("NCMEC")[6] (and through that, to law enforcement for follow-up or tracking where appropriate).

5. **How does TikTok inform parents when a child is exposed to sexual material? How does TikTok inform parents when their child is the target of grooming?**

We respect youth privacy and understand that youth may need additional support to understand online privacy risks and assert their rights. TikTok is committed to offering youth enhanced user education, accessible control over the visibility, access and use of their personally identifiable data, and meaningful measures to prevent third party's unauthorized access to their data (this includes parental access to youth's content consumption data without permission/assent from youth).

We encourage parents to have an open dialogue with their teens and have created a Guardian's Guide[7] to help facilitate these conversations. It also outlines an overview of TikTok and the many tools and controls we've built into the product to keep teens and our broader community safe, as well as general information on common internet safety concerns. It also explains TikTok's privacy policy and safety tools, including our Family Pairing feature, which lets parents link their TikTok account to their teen's account to enable a variety of content, privacy, and well-being settings.

6. **Despite having more than 1 billion active users on your platform, TikTok only launched tools for identifying potential child sexual abuse and grooming within the last month. What is your company's plan to develop a truly comprehensive underage threat detection and prevention strategy?**

This is an inaccurate representation of TikTok's efforts. While TikTok is a young company compared to our peers, we have employed human and machine-based moderation tools like photo identification technologies, in alignment with industry standards, to identify and remove exploitative content for many years. In addition, we filter red-flag language and share information with NCMEC about situations that may indicate grooming behavior, according to their policies and industry norms, and have shared information about our efforts publicly in the

[6] https://www.missingkids.org/home
[7] https://www.tiktok.com/safety/en/guardians-guide/

U.S. since at least early 2020[8], and have been a member of the WeProtect Global Alliance since 2020[9], the Technology Coalition since 2021[10]. We received broad praise for our efforts[11] from many leading organizations as far back as 2021.

7. **Several Trust and Safety policy leaders within TikTok testified that they have been asked to "be lenient on creators with more than 5 million followers," indicating that TikTok applies different moderation standards based on the popularity of certain accounts. Can you confirm that TikTok will apply the same content moderation standards to all accounts, regardless of the size of their audience?**

Our Community Guidelines[12] apply equally to all content and we're committed to enforcing them fairly, consistently, and equitably. Higher follower counts do not lead to more lenient moderation.

8. **A 2022 report from Forbes highlighted issues with TikTok's "Only Me" feature, which offers a pathway for predators to access CSAM on TikTok by hosting the images and videos on a singular account and sharing the account login details amongst other predators. What has TikTok done to close this loophole? Do you scan materials held in the Only Me folder for potential CSAM? Do you scan materials posted publicly for potential CSAM?**

Whenever a video is uploaded to the TikTok platform -- including private videos that are viewable only to the account holder -- the content is run through our automated moderation process. We also proactively scan for user password sharing behavior to identify and remove violative accounts. Additionally, search terms related to "post2private" are blocked, and users with "post2private" related bios, handles, or usernames are taken down.

9. **TikTok is used by many unlawful organizations to inform would-be criminals on methods for circumventing the American legal system. Last week, a CBS 60 Minutes report interviewed dozens of Chinese immigrants who credited TikTok for showing the specific steps to enter this country illegally, including the location of a four-foot hole in the fence near San Diego. Reporters "witnessed nearly 600 migrants—adults and children—pass through the gap and onto U.S. soil, unchecked . . . The migrants knew about the hole because of TikTok. Posts on the app reviewed by '60 Minutes' featured step-by-step instructions for hiring smugglers and detailed directions to the border gap." What is TikTok doing to shut down these types of videos, and other videos that encourage individuals to break the law?**

[8] https://newsroom.tiktok.com/en-us/protecting-against-exploitative-content
[9] https://newsroom.tiktok.com/en-us/tiktok-joins-weprotect-global-alliance
[10] https://newsroom.tiktok.com/en-us/tiktok-joins-the-technology-coalition/
[11] https://newsroom.tiktok.com/en-us/strengthening-privacy-and-safety-for-youth
[12] https://www.tiktok.com/community-guidelines/en/

This type of content would violate TikTok's Community Guidelines. TikTok prohibits content that seeks to promote or facilitate criminal activities, including human smuggling. We do not allow content relating to:

- facilitating or promoting human trafficking and human smuggling activities
- requesting support for being smuggled illegally into a different country
- instructional content on how to illegally cross the border

When we identify content that violates our policies, it is removed from the platform. Our policies relating to human exploitation allow certain content, such as content that expressing a desire to migrate to another country, or showing a migrant's journey (as long as it does not explicitly show the involvement of smugglers in their journey or provide instructions on how to illegally cross the border). We are committed to upholding human rights and preventing the platform from being misused to enable any exploitative activities. We do not allow human exploitation on our platform, including trafficking, smuggling, forced labor, or underage marriage.

We recognize our responsibility to help maintain a safe and welcoming environment for our community. Our Law Enforcement Response Team discloses relevant user data in response to valid legal requests from law enforcement agencies, including DHS. TikTok may also share content or account information directly with law enforcement in the absence of a request when it believes in good faith that there is an emergency involving imminent harm or risk of death or serious physical injury to a person. We also work with third-party intelligence firms to bolster our defenses and make reports to law enforcement as appropriate.

TikTok's Responses to Questions for the Record from Senator Cruz
Submitted March 7, 2024

1. **In the last two years, has an employee or commissioner of the Federal Trade Commission (FTC) requested to evaluate or evaluated your data used for training Large Language Models or algorithms or the sources of such data for bias, discrimination, or misinformation?**

Yes, insofar as TikTok, along with several other companies, received requests for information from the FTC under its 6(b) authority regarding algorithms and bias.

2. **In the last two years, has an employee or commissioner of the FTC sought details regarding your company's measures related to filtering or blocking inputs and outputs of a Large Language Model or algorithms.**
   a. **If yes, has the FTC attempted to coerce or otherwise request you to implement input/output filtering in order to allegedly comply with federal law?**

No. The FTC has not made such a request.

3. **In the last two years, has an employee or commissioner of the Federal Trade Commission sought to evaluate your company's use of measures, including "prebunking" or "debunking", designed to counteract so called "online misinformation"?**

Yes, insofar as TikTok, along with several other companies, received requests for information from the FTC under its 6(b) authority regarding measures to address misleading, deceptive, or fraudulent ads.

4. **In June 2022, the FTC released a report titled "Combatting Online Harms Through Innovation." In this report, the FTC discussed how the deployment of AI tools intended to detect or otherwise address harmful online content is accelerating but may never be appropriate as an alternative to human judgment.**

   a. **In the context of protecting children from online harms to what extent does your company rely on automated tools to detect online harm vs. human review? Please be specific.**

TikTok removes violative content proactively using both automated and manual processes, as well as removing violative content reported by users. Videos uploaded to TikTok are initially reviewed by TikTok's automated moderation technology, which aims to identify content that violates the Community Guidelines before it is distributed across the TikTok platform and displayed to users. These automated technology systems look at a variety of signals across content, including keywords, images, titles, descriptions, and audio, and continuously learn and

adapt based on the data in each video and the moderation decisions that TikTok's human moderators ultimately make based on TikTok's Community Guidelines and related policies. If TikTok's machine-based filters identify a potential violation, the automated moderation system will either pass it on to TikTok's safety teams for further review or remove it automatically. Automatic removal is applied if there is a high degree of confidence that the content violates the Guidelines and where violations are most clear-cut, such as nudity or youth safety. Automatic removals are subject to the user's ability to appeal that determination.

When TikTok's automated moderation systems identify potentially problematic content but cannot make an automated decision to remove it, they send the content to TikTok's safety teams for further review. To support this work, TikTok has developed technology that can identify risky or suspicious items in video frames, so that content moderators can carefully review the video and the context in which it appears. This technology improves the efficiency of TikTok's moderators by helping them more adeptly identify violative images or objects, quickly recognize violations, and make decisions accordingly.

b. What benefits can AI provide to helping detect and/or stop harmful content to children online?

As mentioned in our response to Question 4(a), we use both automated moderation technology and content moderators to identify content that violates our Community Guidelines. We strive to remove violative content as soon as possible, and before it's reported to us by our community. In our latest transparency report[1] covering Q3 2023, our proactive removal rate for violative content was 96.1%, and 90.6% within 24 hours of posting. We continue to improve our content moderation systems to more effectively remove violative content at scale.

c. What does a human reviewer provide that an AI or automated tool cannot? Will we always need some measure of human review in assessing online harms to children?

In order to support fair and consistent review of potentially violative content, we deploy a combination of automated technology and skilled human moderators who can take into account additional context and nuance which may not always be picked up by technology. Human review also helps improve our automated moderation systems by providing feedback for the underlying machine learning models to strengthen future detection capabilities. This continuous improvement helps to reduce the volume of potentially distressing videos that moderators view and enables them to focus more on content that requires a greater understanding of context and nuance. The responsibilities of content moderators include:

- **Reviewing content flagged by technology:** When our automated moderation systems identify potentially problematic content but cannot make an automated decision to remove it, they send the content to our safety teams for further review. To support this work, we've developed technology that can identify risky or suspicious items – for example, weapons – in video frames, so that content moderators can carefully review the

[1] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/

video and the context in which it appears. This technology improves the efficiency of our moderators by helping them more adeptly identify violative images or objects, quickly recognize violations, and make decisions accordingly.

- **Reviewing reports from our community:** We offer our community easy-to-use in-app and online reporting tools so they can flag any content or account they feel is in violation of our Community Guidelines. While these reports are important, the vast majority of removed content is identified proactively before it receives any views or is reported to us.

- **Reviewing popular content:** Harmful content has the potential to rapidly gain popularity and pose a threat to our community. In order to reduce this risk, our automated moderation systems may send videos with a high number of views to our content moderators for further review against our Community Guidelines.

- **Assessing appeals:** If someone disagrees with our decision to remove their content or account, they can file an appeal for reconsideration. These appeals will be sent to content moderators to decide if the content should be allowed back onto the platform or the account reinstated.

d. The FTC has sent mixed signals in its enforcement of COPPA. While the Commission emphasizes not over relying on use of automated tools or AI, they have nonetheless found liability for using human review as alternative signaling overreliance on automated tools. What improvements, if any, should Congress make to clarify the legal tension between use of automated detection tools vs. human review?

We believe that Congress can clarify the legal tension between the use of automated detection tools and human review by focusing on protecting against harmful outcomes, being technology-neutral about specific measures that organizations adopt to prevent these outcomes, and establishing voluntary safe harbors for organizations that implement such measures to mitigate harmful outcomes. Legislation that focuses on protecting against harmful outcomes without being overly prescriptive about the precise measures that organizations must adopt would encourage organizations to develop and innovate new and better automated and human-driven processes that improve harm reduction and prevention. A potential drawback of specifying the tools and technological parameters that organizations must deploy is that these specifications may become obsolete over time as bad actors use increasingly advanced and sophisticated means to achieve their objectives. Congress should consider establishing voluntary safe harbor programs for organizations to innovate and develop technology designed to protect children with appropriate safeguards. By supporting principles-based, harm-focused and technology-neutral online safety laws that include scientifically-based safe harbors, Congress can provide greater regulatory guidance to law enforcement authorities, online service providers, parents and children.

5. In 2021, Congress directed the FTC to research and report on how AI can be used positively to detect and combat fraudulent or deceptive content online. Rather than viewing AI as a potential solution to our online woes, the FTC instead issued a report that read more like an indictment of the technology.

a. Please explain whether, in your view, AI can be used to positively detect and combat fraudulent or deceptive content, including the recent use of deepfakes or other scams to harm consumers.

AI can be used to improve positively detecting and combating fraudulent or deceptive content. AI technologies designed for online fraud detection offer a number of advantages that, when deployed in conjunction with human moderators, can be designed to more effectively reduce harm to consumers from deepfakes or other scams at scale in a number of ways:

1. Models can be trained on datasets that include existing examples of fraudulent or deceptive content in order to learn how to better detect existing or novel harmful attempts and by identifying trends and patterns of coordinated fraudulent behavior.
2. Models can augment human review by quickly identifying potentially fraudulent or deceptive content at a large scale.
3. Models can be used to intake and process reports from users to ensure completeness and reduce wait times for better customer service.

b. Has the FTC ever consulted with your company to learn how your company deploys AI to better detect and combat fraudulent or deceptive content? Has the DOJ? How about the Federal Elections Commission?

Yes, insofar as TikTok, along with several other companies, received requests for information from the FTC under its 6(b) authority regarding use of AI to detect misleading, deceptive, or fraudulent ads.

c. How can Congress empower agencies to use AI positively for the protection of American consumers from fraudulent or deceptive content?

Congress is taking prudent steps to educate itself on the potential benefits and harms of AI development across many sectors, like healthcare, education, and social media. Carefully balancing responsible innovation by technology developers along with safety, security, and privacy for consumers is paramount. Working together with the Administration and other global policymaking efforts, Congress can allocate appropriations in support of agency acquisitions of specified AI products and services that are designed to help consumers identify, report, and prevent the spread of fraudulent or deceptive content.

6. **Please provide a description of your company's policy regarding the sale or transfer of the data of American users collected on your platform to a third party, including data brokers.**

TikTok does not sell personal information. The TikTok U.S. Data Security Inc. ("USDS") vendor program restricts transfer to and/or sharing of U.S. user data to third parties. USDS has established a process to vet any third party who will: (1) have to access, process, modify and/or store TikTok U.S. user data; or (2) provide services to USDS that support operational and daily functions. If the third party fails the compliance assessment, USDS does not contract with the third party. If the third party passes the compliance assessment and no sensitive company information will be shared with the third party, USDS may proceed with contracting. If the third party passes the compliance assessment and sensitive company information will be shared, the third party is then further reviewed from a security risk standpoint as outlined in the following paragraph.

USDS completes a third-party risk and security assessments for any third party that receives sensitive company information, including personal information. To evaluate third party security risk, USDS obtains third party engagement information from completed submission forms and correspondence with associated company business teams. Based upon the security risk profile of the vendor engagement, USDS requests commensurate security documentation from the third party (this may include SOC2 attestations, ISO certifications, policies, penetration and vulnerability reports). USDS reviews this documentation to determine whether the third party meets company security requirements. As necessary, USDS enforces remediation prior to authorizing the vendor.

7. **Has your company ever sold the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the categories of data sold.**

No. We do not sell TikTok user data to governments.

8. **Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the circumstances underlying the basis for such transfer.**

Yes. In response to emergency requests or proactively to prevent emergency situations, we have produced U.S. user data to non-U.S. law enforcement authorities. These requests are not legal orders, but instead emergency related requests from foreign authorities or proactively identified emergency related content that we believe meet a high standard of imminence, credibility and specificity to grant a limited and specific disclosure of user data. In 2023, we disclosed to the following countries: Canada, Mexico, and UK. Examples of emergency nature of requests include: missing minors, child exploitation, violent credible threats, and suicidal missing persons.

9. Has your company ever sold the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and the categories of data sold.

No. We do not sell TikTok user data to governments.

10. Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and categories of data transferred.

Yes. First, in response to emergency requests, we have produced U.S. user data to U.S. law enforcement authorities. These requests are not formal legal process, but instead emergency related requests from U.S. authorities that we believe meet a high standard of imminence, credibility and specificity to grant a limited and specific disclosure of user data. These disclosures are made to assist in either identifying or locating an individual to prevent emergencies.

Examples of the emergency nature of these disclosures include: crimes of violence, school shootings, national security/ terrorism, suicide/self-harm, missing persons, kidnappings, mass shooting/ casualty events, and child exploitation. These requests typically include basic subscriber information and IP log-in data.

In 2023, the U.S. Federal government agencies that have received U.S. user data under this policy and process are:

- FBI
- DHS (HSI, CBP)
- U.S. Marshals Service
- U.S. Secret Service
- U.S. Capitol Police
- U.S. Treasury Inspector General for Tax Administration
- USNCB Interpol Washington

Second, we also have proactively disclosed U.S. user data to U.S. law enforcement authorities to report potential emergency situations or provide evidence of potential serious criminal conduct. These data disclosures are done in response to internally identified content that either a) meet a high emergency standard of imminence, credibility and specificity or b) contains evidence of potential serious criminal conduct.

Examples of the emergency nature of these disclosures include: suicide/self-harm, crimes of violence, school shootings, national security/ terrorism, mass shooting/ casualty event, child exploitation and kidnapping. These requests typically include: basic subscriber information, IP log-in data, user content, and metadata about the user content. User content and metadata are produced here to provide the evidence of the potential emergency.

Examples of the potential serious criminal conduct included in the below disclosures: human smuggling or trafficking, firearms trafficking, and drug trafficking. These requests typically include basic subscriber information, IP log-in data. user content, and metadata about the user content. User content and metadata are produced here to provide the evidence of the potential serious criminal conduct.

In 2023, we produced U.S. user data according to these policies and processes to the following U.S. federal government agencies:

- Potential Emergencies:
  - FBI
  - USNCB Interpol Washington
- Potential Serious Criminal Conduct:
  - ATF
  - DEA
  - FBI

11. Does your company have a policy to restrict third party use and/or transfer of data collected from users on your platform? Please be specific, including how you enforce such restrictions and whether such restrictions prohibit the sale or transfer of such data to a government agency, including a foreign government agency.

TikTok does not sell personal information. The TikTok U.S. Data Security Inc. ("USDS") vendor program restricts transfer to and/or sharing of U.S. user data to third parties. USDS has established a process to vet any third party who will: (1) have to access, process, modify and/or store TikTok U.S. user data; or (2) provide services to USDS that support operational and daily functions. If the third party fails the compliance assessment, USDS does not contract with the third party. If the third party passes the compliance assessment and no sensitive company information will be shared with the third party, USDS may proceed with contracting. If the third party passes the compliance assessment and sensitive company information will be shared, the third party is then further reviewed from a security risk standpoint as outlined in the following paragraph.

USDS completes a third-party risk and security assessments for any third party that receives sensitive company information, including personal information. To evaluate third party security risk, USDS obtains third party engagement information from completed submission forms and correspondence with associated company business teams. Based upon the security risk profile of the vendor engagement, USDS requests commensurate security documentation from the third party (this may include SOC2 attestations, ISO certifications, policies, penetration and vulnerability reports). USDS reviews this documentation to determine whether the third party meets company security requirements. As necessary, USDS enforces remediation prior to authorizing the vendor.

12. Between July 4, 2023 and July 14, 2023, was your company contacted by any employee of or contractor for any of the following agencies? Please answer "yes" or "no" for each agency and, if "yes," provide the date(s) of contact and the name(s) of the agency employees or contractors that contacted your company.

a. U.S. Department of Health and Human Services (HHS)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

b. National Institute of Allergy and Infectious Diseases (NIAID)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

c. Centers for Disease Control and Prevention (CDC)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

d. U.S. Food and Drug Administration (FDA)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

e. The National Institutes of Health (NIH)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

f. U.S. Department of Homeland Security (DHS)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

g. DHS Cybersecurity and Infrastructure Security Agency (CISA)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

h. U.S. Census Bureau

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

i. Federal Bureau of Investigation (FBI)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

j. U.S. Department of Justice (DOJ)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

k. The White House Executive Office of the President (EOP)

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

l. U.S. Department of State

Other than any valid legal process, we have not to date as part of a limited informal review identified any communication by any employee of or contractor of this agency during the time period of July 4, 2023 to July 14, 2023. We may be unaware of communications if they took place between an agency and an employee and it was not reported to the company.

13. Is it your company's policy to prevent children under 13 from using your social media app(s) or creating an account?

In line with the FTC's guidance, under our policy in the U.S., users age 13 and older may access TikTok's 13+ experience, but for users under 13, we provide a separate under 13 experience which has additional safeguards and privacy protections designed specifically for children under 13.

14. In your view, would it be appropriate for school-aged children to spend time on or access your company's social media app(s) during class?

Every teen is different, and so is every family, but for all of us there are times when it's important to be uninterrupted. This is why caregivers are able to use Family Pairing to customize the daily screen time limit for their teen – including choosing different time limits depending on the day of the week – giving families choices to match screen time to school schedules, holiday breaks, or family travel. TikTok has also enabled parents to set a schedule to mute notifications for their teen.

TikTok believes that schools and teachers are best positioned to make determinations as to their classroom curriculums.

15. As a parent, would you be concerned if your child were able to access your company's social media app(s) during class via a school network or device?

Every teen is different, and so is every family, but for all of us there are times when it's important to be uninterrupted. This is why caregivers are able to use Family Pairing to customize the daily screen time limit for their teen – including choosing different time limits depending on the day of the week – giving families choices to match screen time to school schedules, holiday breaks, or family travel. TikTok has also enabled parents to set a schedule to mute notifications for their teen.

TikTok believes that schools and teachers are best positioned to make determinations as to their classroom curriculums.

16. In your view, should elementary and secondary schools block students' access to your company's social media app(s) on school networks and devices?

Every teen is different, and so is every family, but for all of us there are times when it's important to be uninterrupted. This is why caregivers are able to use Family Pairing to customize the daily screen time limit for their teen – including choosing different time limits depending on the day of the week – giving families choices to match screen time to school schedules, holiday breaks, or family travel. TikTok has also enabled parents to set a schedule to mute notifications for their teen.

TikTok believes that schools and teachers are best positioned to make determinations as to their classroom curriculums.

17. Do you think that school buses equipped with Wi-Fi should allow children to access your company's social media app(s) via a school bus Wi-Fi network during their rides to and from school?

Every teen is different, and so is every family, but for all of us there are times when it's important to be uninterrupted. This is why caregivers are able to use Family Pairing to customize the daily screen time limit for their teen – including choosing different time limits depending on the day of the week – giving families choices to match screen time to school schedules, holiday breaks, or family travel. TikTok has also enabled parents to set a schedule to mute notifications for their teen.

TikTok believes that school districts are best positioned to make determinations as to what content should be available from school bus Wi-Fi networks.

18. As a parent, do you think it is important to supervise your children's internet access?

TikTok encourages parents and guardians to have ongoing dialogue and supervision of their teen's digital journey. TikTok provides a Guardian's Guide to help facilitate conversations between parents and their teens. More information on our approach can be found here[2].

19. As a parent, would you be concerned if your child's school allowed your child to access the internet on an unsupervised basis, such as on your child's bus ride to and from school via the school bus Wi-Fi?

TikTok agrees that it is important for parents to understand how their children's schools engage with technology and what safeguards will be put in place.

20. Do you think Congress should require schools, as a condition of receiving broadband subsidies through the Federal Communications Commission's E-Rate program (which funds broadband for elementary and secondary schools), to block students' access to your company's social media app(s) from school-run networks?

TikTok has not taken a position on the E-Rate program. TikTok has significant educational content on the app, such as our STEM feed[3] or #booktok content[4] that schools may wish to allow students to access.

[2] https://www.tiktok.com/safety/en/guardians-guide/
[3] https://newsroom.tiktok.com/en-us/take-discovery-to-a-new-level-with-stem-feed
[4] https://newsroom.tiktok.com/en-us/world-book-day-booktok-2023

21. Do you support the bipartisan *Eyes on the Board Act of 2023*, S. 3074?

No.

22. Have you, your company, or any foundation associated with you or your company, donated or contributed funding, equipment, or services to any of the following organizations in the last ten years (CY 2013 to CY 2023)?
- **a. Education and Libraries Networks Coalition (EdLiNC)**
- **b. Open Technology Institute**
- **c. Consortium for School Networking (COSN)**
- **d. Funds For Learning**
- **e. State Educational Technology Directors Association (SETDA)**
- **f. Schools, Health, and Libraries Broadband Coalition (SHLB)**
- **g. State E-Rate Coordinators' Alliance (SECA)**
- **h. EducationSuperHighway**
- **i. All4Ed**
- **j. Public Knowledge**
- **k. Fight for the Future**
- **l. Free Press**
- **m. Electronic Frontier Foundation**
- **n. Benton Foundation or Benton Institute for Broadband & Society**
- **o. Electronic Privacy Information Center**

TikTok has contributed funding to Public Knowledge.

23. For each such donation or contribution described in the prior question, please detail (1) the type of donation or contribution, such as financial donation, goods or equipment, services, etc.; (2) who made the donation or contribution; (3) the recipient organization; (4) the year the donation or contribution was made; and (5) the total value of that donation or contribution.

In 2021, TikTok Inc. contributed $50,000 to support Public Knowledge's IP3 Awards.

In 2022, TikTok Inc. contributed $50,000 to support Public Knowledge's IP3 Awards.

In 2023, TikTok Inc. pledged $50,000 to support Public Knowledge's IP3 Awards and another $50,000 to support Public Knowledge's Sherwin Lee Siy Memorial Fund, which funds Public Knowledge's training program for diverse early career public interest advocates.

24. China's 2017 national intelligence law states: "All organizations and citizens shall support, assist, and cooperate with national intelligence efforts in accordance with the law, and shall protect national intelligence work secrets they are aware of."

a. You worked for Chinese tech giant Xiaomi from 2015 to 2021. Yes or no: As an employee of Xiaomi, were you covered by China's national intelligence law following its 2017 passage?

It is our understanding that any business that operates in China, including many other global businesses, is subject to Chinese law.

b. Yes or no: Is Xiaomi as an organization covered by the 2017 national intelligence law?

It is our understanding that any business that operates in China, including many other global businesses, is subject to Chinese law.

c. You worked as ByteDance's chief financial officer in 2021. Yes or no: As an employee of ByteDance, were you covered by China's national intelligence law?

ByteDance is obligated to comply with China's laws insofar as they relate to the ByteDance business just as it is obligated to comply with the laws of the United States. Other U.S. companies doing business in China would likewise be required to comply with Chinese laws insofar as they are doing business in China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

d. Yes or no: Is ByteDance as an organization covered by the 2017 national intelligence law?

ByteDance is obligated to comply with China's laws insofar as they relate to the ByteDance business just as it is obligated to comply with the laws of the United States. Other U.S. companies doing business in China would likewise be required to comply with Chinese laws insofar as they are doing business in China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

e. In March 2023, you testified before Congress that you report to the CEO of ByteDance, Liang Rubo. Yes or no: Is Liang Rubo covered by the 2017 national intelligence law?

ByteDance is obligated to comply with China's laws insofar as they relate to the ByteDance business just as it is obligated to comply with the laws of the United States. Other U.S. companies doing business in China would likewise be required to comply with Chinese laws insofar as they are doing business in China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

f. Yes or no: As CEO of TikTok—a role in which you report to the CEO of ByteDance—are you covered by the 2017 national intelligence law?

TikTok does not offer the TikTok app for download in mainland China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

g. Yes or no: Is TikTok as an organization covered by the 2017 national intelligence law?

TikTok does not offer the TikTok app for download in mainland China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

h. How many of ByteDance's employees are covered by the 2017 national intelligence law?

TikTok does not offer the TikTok app for download in mainland China.

ByteDance is obligated to comply with China's laws insofar as they relate to the ByteDance business just as it is obligated to comply with the laws of the United States. Other U.S. companies doing business in China would likewise be required to comply with Chinese laws insofar as they are doing business in China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

i. How many of TikTok's employees are covered by the 2017 national intelligence law?

TikTok does not offer the TikTok app for download in mainland China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

25. Is TikTok subject to the Cybersecurity Law of the PRC, and/or is TikTok subject to the regulations put forward by Cyberspace Administration of China (CAC), their 2021 Regulations on the Management of Network Product Security Vulnerabilities" (RMSV)? Please be specific.

a. If yes, what are TikTok's obligations to disclose the cyber vulnerabilities in its application to America's government or America's TikTok users? Are they different from TikTok's obligations to the PRC? Does TikTok commit to notify the U.S. government and consumers if and when it discovers a cyber vulnerability concurrently with the PRC should that situation arise?

TikTok does not make the app available for download in mainland China. Additionally, we've taken steps to limit to only USDS personnel access to protected U.S. user data in the Oracle environment, unless authorization is given by USDS pursuant to limited exceptions. Traffic from the Oracle Cloud now goes through Oracle controlled gateways to prevent protected U.S. TikTok user data from being transferred to or accessed by employees of TikTok or ByteDance. TikTok believes that these cutting-edge measures are unprecedented among our peer group and will provide many layers of protection for U.S. user data.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

26. Last year, it was reported that TikTok and ByteDance employees have access to a tool internally known as "heating," which can be used to manually boost videos in TikTok's "For You" feed—resulting in supercharged reach to the platform's 1.1 billion users.

a. How many TikTok employees currently have the ability to "heat" videos?

Only a small number of people, based in the U.S., have the ability to approve content for promotion in the U.S., and that content makes up approximately 0.002% of videos in For You feeds.

b. According to an internal document called "TikTok Heating Policy," heating is often used for commercial purposes, such as helping influencers or brands go viral. However, it can also be used to "push important information." How does TikTok define "important information" in its Heating Policy?

We are uncertain which specific document is referenced by the description "TikTok Heating Policy".

The TikTok app may use "in-app notifications" for users when they have new messages, interactions, and more. Additionally, in-app notifications may be used to share information such as to alert users to new features on the app or to alert creators to new commercial programs. Users have the ability to customize the kinds of in-app notifications that they receive, and can set a personalized schedule to mute in-app notifications.

c. Yes or no: Has content posted by a Chinese state media outlet ever been heated on TikTok?

i. If "yes," identify (1) the outlet, (2) a description of the piece of content, (3) the date(s) for which the content was heated, and (4) the

number of views that the piece of content received in the United States. Provide this information for each piece of heated content.

TikTok has policies that prohibit personnel from taking any action in the course of their work for or with TikTok to advance the political agenda of a third party through the promotion, recommendation, moderation, or filtering of content.

d. Yes or no: Has content posted by a Russian state media outlet ever been heated on TikTok?

i. If "yes," identify (1) the outlet, (2) a description of the piece of content, (3) the date(s) for which the content was heated, and (4) the number of views that the piece of content received in the United States. Provide this information for each piece of heated content.

TikTok has policies that prohibit personnel from taking any action in the course of their work for or with TikTok to advance the political agenda of a third party through the promotion, recommendation, moderation, or filtering of content.

e. Yes or no: Has TikTok ever received a request from ByteDance to heat content?

TikTok may promote or "heat" specific content (including, e.g., promoting the video of an artist who will be hosting a concert on TikTok Live) in line with company content policies to support the inclusion of diverse and high-quality content on the platform. A content operations team will review heating requests submitted by a limited number of cross-functional partners with access to the heating request, and the Content Operations team will either approve or reject the request based on their assessment of whether it follows the platform's best practices in support of content diversity and quality (including, e.g., being engaging and meaningful and focusing on timely/relevant content) and business objectives. A USDS Trust and Safety team also reviews Heating requests to help ensure no content violating its policies will be promoted using the Heating function. Even if the request is approved, increasing visibility or video views ("VV") is not guaranteed as the recommendation system will not recommend low quality content (e.g, reposted or irrelevant content). Heating impacts less than 1% of VV in the U.S. We have guidelines for the process of human-curation of content known as "heating", but do not publicly disclose those guidelines in order to prevent bad actors from attempting to manipulate our systems.

27. You and TikTok U.S. Data Security (USDS) staff have said that TikTok has "given" Oracle full access to TikTok's source code to review as part of TikTok's "Project Texas" data security plan.

a. How many lines of source code does TikTok have?

Similar to other feature rich applications, TikTok is made up of millions of lines of code.

b. How will Oracle inspect every line of TikTok source code to ensure there are no backdoors?

We are committed to protecting TikTok systems from unauthorized access to protected U.S. user data. Oracle has both the access to the source code and the environment to do the source code analysis as they see fit. It includes a combination of human and automated analysis. Oracle keeps the details of their approach confidential to maintain the integrity of their tools, techniques, and analysis. Additionally, we plan to announce the selection of another independent security inspector who, in addition to Oracle, will have access to our source code and systems, further testing the security and integrity of TikTok's platform and its software.

c. When did TikTok "give" Oracle its source code? Provide an exact date.

Oracle began reviewing the TikTok app source code in January 2023. This review is on-going, and we plan to announce the selection of another independent security inspector who, in addition to Oracle, will have access to our source code and systems, testing the security and integrity of TikTok's platform and its software.

d. By what mechanism was the source code provided?

We have opened a Dedicated Transparency Center ("DTC") in Maryland where Oracle has access to review TikTok source code. As of January 2024, we have opened an additional DTC in Colorado, and we expect to have two fully operational DTCs in the United Kingdom and Australia soon. There are independent experts reviewing TikTok code to find and eliminate security vulnerabilities.

e. Did TikTok give Oracle a copy of its source code, or was Oracle granted access to TikTok's codebase?

TikTok and related source code must go through the Software Assurance process in order to run in the secure Oracle environment. It is first loaded into an instance of the Oracle Cloud called the Secure Computing Environment; from there, it is available for Oracle to conduct automated and manual analysis prior to it being compiled and deployed by USDS. Only source code that comes through this analysis is allowed to run in the secure Oracle Cloud environment. Code for the mobile app, which is deployed to the app stores by global teams, is also available to Oracle to review.

f. Will Oracle have access to TikTok's codebase on an ongoing basis? If yes, please describe. If no, please explain why not.

Yes. TikTok and related source code must go through the Software Assurance process in order to run in the secure Oracle environment. It is first loaded into an instance of the Oracle Cloud called the Secure Computing Environment; from there, it is available for Oracle to conduct automated

and manual analysis prior to it being compiled and deployed by USDS. Only source code that comes through this analysis is allowed to run in the secure Oracle Cloud environment. Code for the mobile app, which is deployed to the app stores by global teams, is also available to Oracle to review.

28. TikTok's divestment from ByteDance is one possible solution to ongoing national security concerns related to TikTok's Chinese ownership. In the event that TikTok is sold to an American owner, how will you ensure that during the sensitive post-sale period all American information is processed within American jurisdictions and following American privacy laws as the Transition Services Agreement (TSA) is negotiated?

TikTok believes that U.S. national security concerns can be addressed through mitigation measures, such as data privacy safeguards, that have been applied for years under similar circumstances. The company has already devoted substantial resources over the course of many years to adopt such measures and has made itself open to intensive scrutiny from stakeholders as well. For these reasons, TikTok believes divestiture is not necessary and would violate fundamental rights of the company and its users.

29. During our meeting, you mentioned that TikTok doesn't use end-to-end encryption so you can monitor/moderate certain content on the platform.

a. Does TikTok use end-to-end encryption for adult users or is it a platform-wide policy – i.e. for adult users and minor users – not to use end-to-end encryption?

No. TikTok does not currently use end-to-end encryption (E2EE) for direct messages.

b. As I mentioned during the hearing, I am concerned about TikTok's ties to the Chinese Communist Party. You noted that your technology (like all technology) is not completely free of a vulnerability of hackers. Yes or no: Doesn't using encryption for TikTok users increase TikTok's cybersecurity vulnerabilities, including the vulnerability of U.S. user data from hackers or other unauthorized access?

No, we do not believe the use of encryption increases TikTok's cybersecurity vulnerabilities. TikTok uses industry-standard encryption to protect sensitive user data. We encrypt sensitive user data in transit and at rest. Data can only be decrypted with a key that is generated and managed by our key management service, which is operated by USDS.

30. As you are aware, Members of Congress on both sides of the aisle have serious concerns with TikTok's ties to the Chinese Communist Party. During our meeting, we discussed at length about who has access to TikTok's U.S. user data. You noted during our meeting that "protected data" is stored on U.S. soil and run through Oracle and TikTok's U.S. Data Security (USDS). You firmly stated that no one

outside of either Oracle or USDS can obtain access to this data. This assurance, of course, only applies to "protected data".

a. In reviewing TikTok's privacy policy, I did not find a definition for data that is "not protected". Please provide TikTok's definition of "not protected" data.

"Protected data" is a concept used by CFIUS to distinguish between types of data that require protection under U.S. national security policy and other types of data, such as public information, that do not. Under Project Texas, "protected data" of U.S. users is afforded stricter protection. Protected data broadly means personal information collected from a TikTok U.S. user. subject to exceptions, protected data include the following categories of data: user data, such as email and birthdate; non-public user content, such as private videos and direct messages; behavioral data, such as user interaction with content including likes and favorites; data inputs to TikTok's recommendation engine, such as video completion and video viewing time; and device and network data, such as IP address and device model.

TikTok has been clear that there are certain, limited exceptions to the definition of protected data. These exceptions are in place to help ensure interoperability of TikTok as a global platform and were determined as part of TikTok's extensive, multi-year discussions with CFIUS that have occurred under two Administrations.

These necessary exceptions to protected data include: business metrics such as daily active user stats; data of certain creators pursuant to an agreement; interoperability data such as data needed to apply a user's privacy settings globally; and e-commerce data such as shipping information.

b. TikTok's most recent Privacy Policy includes "exceptions to protected data" policies including data "that allow TikTok to continue operating a business and as a global platform, including public data such as public videos." Please define "public data".

Public data is data that is generally accessible to other users of the TikTok platform, such as publicly posted videos.

c. Does setting my TikTok account to "public" rather than "private" expand the categories of data accessible outside the United States? If so, which data categories?

By changing your account settings to public, your public content will then available to TikTok users around the world to view and interact with where TikTok is available. You will still have the ability to adjust public/private settings for each individual video you post.

d. Please provide the most recent data on the number of TikTok user accounts set to public. Please provide the most recent data on the number of user accounts on TikTok set to private.

As a privately held company, TikTok does not disclose detailed statistics about its users or adoption of specific features.

e. Please provide the most recent data on the number of TikTok minor user accounts who have their accounts set to public.

As a privately held company, TikTok does not disclose detailed statistics about its users or adoption of specific features; however, all accounts of 13-15 year olds are set to private by default.

f. Please provide the most recent data on the number of TikTok minor user accounts who have their accounts set to private.

As a privately held company, TikTok does not disclose detailed statistics about its users or adoption of specific features; however, all accounts of 13-15 year olds are set to private by default.

g. Does TikTok apply its policies regarding access to "public data" outside the United States equally to both adult and minor TikTok user accounts who are set to public rather than private? In other words, do minor TikTok user accounts set to public have more or the same data protections as adult TikTok user accounts set to public?

We have prioritized the privacy of younger users and made substantial product improvements to help parents and minors navigate the platform safely, and continue to focus on improving privacy safeguards for minors that use TikTok. As of January 2021, all under 16 accounts are private by default. Even if an account registered to a teen under 16 chooses to change their private account setting, others are still unable to download, Duet, or Stitch their content, and the ability to "suggest your account to others" is turned off by default. In addition, their content will not be recommended in the For You feed to people they do not know. Direct messaging is not available for teens under 16.

h. A common TikTok user goal is to go "viral" which is more likely to happen if your account is public. If more TikTok user accounts are public with significantly more data accessible outside the U.S., then why shouldn't we view this as a huge loophole undermining your reliance on data security provided by Project Texas?

Many other apps and platforms enable public sharing of content. On TikTok, users can choose whether to have a private account or a public account and can make updates through in-app settings.

31. In December 2022, news broke that ByteDance fired four employees, including two employees based in China, who improperly accessed the personal data of two journalists on the platform. This data included IP addresses, which can provide

information about a user's location. TikTok's Privacy Policy states that "As of January 2023" (just a few weeks after this story broke) that all access to new U.S. user protected data is exclusively controlled by the USDS.

a. Please provide a list of the categories of personal data that these fired employees based in China were able to access.

Based on our internal investigation of the facts surrounding the incident, the two China-based former employees did not obtain access to TikTok's internal databases that contain TikTok user data in connection with this conduct.

b. Was any of the personal data accessed by the employees based in China, data that TikTok considers "not protected"? If so, which data categories?

Based on our internal investigation of the facts surrounding the incident, the two China-based former employees did not obtain access to TikTok's internal databases that contain TikTok user data in connection with this conduct.

c. Please describe the actual steps TikTok took in the span of a couple of weeks to close off all access to new U.S. user protected data from employees outside USDS.

The process of separating individual production systems and databases began July 2022 through January 2023 when final account transitions were completed.

d. Can employees based in China access U.S. user data that is "not protected"? If so, how many employees are based in China with access to this data?

Public videos and related data are accessible worldwide, just like content on the internet generally.

e. How many employees not based in China or the United States have access to U.S. user data that is "not protected"?

Public videos and related data are accessible worldwide, just like content on the internet generally.

f. As noted above, TikTok's policies state that USDS's exclusive control over access to user data only applies to "new U.S. user protected data." What about existing U.S. user protected data? Does any employee or person outside the United States have access to U.S. user protected data that was collected or used by TikTok prior to January 2023?

TikTok continues to delete historical protected U.S. user data from our global data centers, helping to ensure that protected U.S. user data is stored in Oracle's Cloud and controlled by

USDS. Since last spring, our team has completed the first round of deletion from TikTok servers in our global data centers. In order to provide additional assurance and validation of our team's work, we have also begun independent validation efforts to confirm the comprehensiveness of the data deletions.

g. TikTok further notes in its privacy policy that "new protected data is restricted to approved USDS employees". Are all of these employees located in the United States? Are any of them subject to the jurisdiction of the CCP? Please be specific.

USDS employs personnel in the U.S., Australia, and in the United Kingdom.

h. TikTok's policies note that there are some limited exceptions for non-USDS employee access to protected data, including for legal and compliance reasons. How many non-USDS employees have this access? Please provide the employer for each non-USDS employee that has this access. Please provide a list of each country where the non-USDS employees are based.

Approximately 50 non-USDS employees, who are based in the U.S., Germany, England, Ireland, Japan, and Singapore have limited access to protected data on an as-needed basis, including for legal and compliance reasons.

i. Please explain TikTok's position on which legal and compliance reasons warrant an exception to this policy. Does compliance with the laws of a foreign country, including China, cover this exception?

TikTok, as a U.S. company incorporated in the United States, is subject to the laws of the United States. In order to adequately conduct litigation, respond to regulatory enforcement actions, and conduct legal investigations, the TikTok legal team must be able to access and analyze relevant data. This can include both protected U.S. user data, access to which is managed by USDS, and other non-protected data. This exception includes U.S. litigation and regulatory matters and compliance investigations, and it has not been used as a basis to share any protected U.S. user data with Chinese authorities. Additional safeguards are built into the process, to include that data sharing would need to be approved by USDS for each instance.

32. TikTok has stated that as of March 2023, they began the process of deleting U.S. user protected data from the data centers in Virginia and Singapore.

a. Has TikTok completed this deletion process? Please be specific.

TikTok continues to delete historical protected US TikTok user data from our global data centers, helping to ensure that protected US TikTok user data is stored in Oracle's Cloud and controlled by USDS. Since last spring, our team has completed the first round of deletion from TikTok servers in our global data centers. In order to provide additional assurance and validation of our team's work, we have also begun independent validation efforts to confirm the comprehensiveness of the data deletions.

b. Does TikTok have additional data centers other than the ones in Virginia and Singapore that store U.S. user protected data? If so, where? Are these data centers deleting U.S. user protected data?

As TikTok announced in June 2022, U.S. user traffic is today routed to the Oracle cloud and USDS Infrastructure in the United States. Moreover, since January 2023, we have implemented controls to block access to databases in Oracle containing new protected U.S. user data and limit to approved USDS employees unless there is a limited exception. TikTok is in the process of deleting historic protected U.S. user data on non-USDS TikTok systems globally; once that process is complete, TikTok will have taken necessary steps to effectively end all access to protected U.S. user data outside of USDS except under limited circumstances being defined through engagement with CFIUS.

c. Please provide a list of all data centers in the world that store U.S. user data that is "not protected".

At this time, TikTok has primary data centers available and online in the United States, Singapore, Malaysia, and Ireland, and U.S. user data that is not protected is stored in those data centers.

33. In TikTok's privacy policy, you note that TikTok does not collect "precise geolocation" but does collect approximate geolocation information based on a SIM card and/or IP address. TikTok further notes that if you are using a previous version of the app last released in August 2020, you may still be collecting precise geolocation.

a. How many U.S. users are still using a previous version of the app that would allow for the collecting of precise geolocation?

TikTok continues to delete historical protected US TikTok user data from our global data centers, helping to ensure that protected US TikTok user data is stored in Oracle's Cloud and controlled by USDS. Since last spring, our team has completed the first round of deletion from TikTok servers in our global data centers. In order to provide additional assurance and validation of our team's work, we have also begun independent validation efforts to confirm the comprehensiveness of the data deletions.

b. Why does TikTok still collect precise geolocation on older versions of the app? Why not just stop the collection/use of such data consistent with your stated policies?

As TikTok announced in June 2022, U.S. user traffic is today routed to the Oracle cloud and USDS Infrastructure in the United States. Moreover, since January 2023, we have implemented controls to block access to databases in Oracle containing new protected U.S. user data and limit to approved USDS employees unless there is a limited exception. TikTok is in the process of deleting historic protected U.S. user data on non-USDS TikTok systems globally; once that

process is complete, TikTok will have taken necessary steps to effectively end all access to protected U.S. user data outside of USDS except under limited circumstances being defined through engagement with CFIUS.

c. As part of the process of deleting U.S. user protected data note above, is TikTok deleting all U.S. user precise geolocation consistent with its data storage policies or does prior user consent with the older version of the app make this data "not protected"?

At this time, TikTok has primary data centers available and online in the United States, Singapore, Malaysia, and Ireland, and U.S. user data that is not protected is stored in those data centers.

34. Please provide a diagram and explanation of TikTok's data architecture and the data flows with respect to Project Texas. Specifically, please include the following:

a. How major updates or patches to the TikTok software are delivered.

The delivery of updates or patches to TikTok software, all of which are reviewed by USDS, can currently be delivered in these ways:
1. Updates or patches submitted by USDS personnel:

1. Deploy in global/Rest of World environment to ensure it works in that environment
2. Submit a deployment request for Trusted Technology Provider environment
3. Review by USDS personnel
4. Run unit test and integration test in Trusted Technology Provider environment
5. Deploy to Trusted Technology Provider pilot
6. Deploy to Trusted Technology Provider all clusters

2. Updates or patches submitted by non-USDS personnel:

1. Deploy in global/Rest of World environment to ensure it works in that environment
2. Submit a deployment request for Trusted Technology Provider environment
3. Review and approve by USDS personnel
4. Run unit test and integration test in Trusted Technology Provider environment
5. Deploy to Trusted Technology Provider pilot
6. Deploy to Trusted Technology Provider all clusters

b. The origin of TikTok software coding and coding fixes or improvements (i.e. patches, updates, maintenance).

Software coding: As part of our US product release process, source code from our global engineering work force is uploaded to the Oracle Cloud Infrastructure environment where it is available to be inspected by Oracle. The process requires code to go through the review process in order to run in the environment. The process requires all TikTok code to go through Oracle's

review process (including technology or human review). Finally, the process also calls for Oracle to compile the app, and deploy it to the app stores, maintaining chain of custody for assurance.

Coding fixes or improvements: The origin of fixes and improvements will depend on whether it is commercial software, opensource, or internally developed. For internally developed software, the source code for fixes is developed globally, but the Software Assurance process (including being available for Oracle to review) requires source code to be deployed in the Oracle Cloud instance that serves U.S. users.

c. Whether the fixes to the code ever originate from outside of the United States.

Software is developed globally but is subject to the Software Assurance process before it is deployed.

d. How and where diagnostics are run on the American TikTok application.

Diagnostics to ensure cloud resources are operating as expected, or that require access to protected U.S. user data are conducted by USDS.

e. How, when, and where data (both protected data and not protected data, including any diagnostic data or that related to service and maintenance) from American TikTok users of the American TikTok application is transferred or accessed outside of the United States.

Three use cases:

1. Protected data: USDS has personnel in Australia and the UK who provide "follow the sun", 24/7 support and who have approved access to protected U.S. user data and can access the U.S. Oracle cloud from these regions.
2. Limited access protocol: The global team can make requests for data for legal, compliance, or safety reasons under limited scenarios. USDS reviews these requests and either rejects or approves them. If appropriate and approved, USDS will provide relevant information to designated members of the global team.
3. Excepted data: Data is verified in one of the Oracle gateways and synchronized with global systems (e.g. performance metrics, etc).

Public data is data that is generally accessible to other users of the TikTok platform, such as publicly posted videos.

Since January of 2023, new protected U.S. user data has been stored in the Oracle Cloud in an environment controlled by USDS. We've taken steps to limit access to protected U.S. user data in the Oracle environment, unless authorization is given by USDS pursuant to limited exceptions, such as for legal and compliance purposes. Traffic from the Oracle Cloud now goes through

Oracle controlled gateways to prevent protected U.S. user data from being transferred to or accessed by employees of TikTok or ByteDance.

35. Please list all of TikTok's donations or contributions of funding, equipment, and/or services to 501(c) organizations in the last six years (CY 2017 to CY 2023). For each such donation or contribution, please detail (1) the type of donation or contribution, such as, financial donation, goods or equipment, services, etc.; (2) the recipient organization; (3) the year the donation or contribution was made; and (4) the total value of that donation or contribution.

TikTok makes contributions to charitable organizations that align with our Global Impact Themes. In addition to financial contributions, from time to time, TikTok may also provide an in-kind donation in the form of ad credits to charitable organizations. For example, TikTok has donated $100,000 in advertising credits to the Ad Council and $25,000 to Song for Charlie[5] to aid their efforts in spreading awareness around the dangers of fentanyl. Cash and in-kind donations can also be allocated in times of emergencies, including natural disasters, public health emergencies and conflicts. For example, TikTok in 2023 donated $300,000 to the Hawaii Community Foundation as well as $200,000 in ad credits to support Maui after the devastating wildfires.

Some examples of our contributions include a $250,000 donation by TikTok to Rare Impact Fund[6] in connection with World Mental Health Day to sponsor the inaugural Rare Impact Fund Benefit and support the organization's global work and $1 million donation by TikTok to each of the Hispanic Heritage Foundation and Black Girl Ventures[7] in connection with National Small Business Week.

Previously Unanswered Question from *Questions for the Record, Senate Commerce Committee Hearing on "Subcommittee: Protecting Kids Online: Snapchat, TikTok, and YouTube" (October 26, 2021)*

36. Please provide the following:
a. The legal name of each "parent, subsidiary, or other affiliate" which is part of TikTok's "corporate group."

TikTok Inc. is incorporated in California and is the provider of the TikTok platform in the United States. TikTok U.S. Data Security Inc. ("USDS") is a wholly owned subsidiary of TikTok Inc. USDS is incorporated in Delaware, and its main responsibility is to keep the data of TikTok's U.S. users secure.

TikTok Inc. is wholly owned by TikTok LLC, a Delaware limited liability company. TikTok LLC is in turn wholly owned by TikTok Ltd., a Cayman Islands company. TikTok's corporate

[5] https://newsroom.tiktok.com/en-us/national-fentanyl-awareness-day
[6] https://newsroom.tiktok.com/en-us/supporting-our-community-this-world-mental-health-day-and-beyond
[7] https://newsroom.tiktok.com/en-us/tiktok-small-business-week-2023

group consists of TikTok Ltd. and its subsidiaries. Besides the subsidiaries described above, TikTok Ltd. maintains subsidiaries around the world to support its global operations. These subsidiaries include, among others, TikTok Pte. Ltd. (established in Singapore), TikTok Information Technologies UK Ltd. (established in the UK), TikTok Technology Ltd. (established in Ireland), TikTok Technology Canada Inc. (established in Canada), TikTok Australia Pty. Ltd. (established in Australia), and TikTok Hong Kong Ltd. (established in Hong Kong). TikTok Ltd. is wholly owned by Bytedance Ltd., a Cayman Islands company.

b. The location where each "parent, subsidiary, or other affiliate" is headquartered

Los Angeles and Singapore are TikTok's global headquarters.

c. The location where each "parent, subsidiary, or other affiliate" is domiciled, if domicile location differs from where the parent, subsidiary, or other affiliate is headquartered.

Please see the locations listed above. Los Angeles and Singapore are TikTok's global headquarters.

d. For each "parent, subsidiary, or other affiliate," the laws under which they were originally incorporated.

TikTok Inc. is incorporated under the laws of California. TikTok U.S. Data Security Inc. ("USDS") is incorporated under the laws of Delaware.

TikTok LLC is organized under the laws of Delaware. TikTok Ltd. is organized under the laws of the Cayman Islands. TikTok Ltd.'s subsidiaries outside of the U.S. include, among others, TikTok Pte. Ltd. (organized under the laws of Singapore), TikTok Information Technologies UK Ltd. (organized under the laws of England and Wales), TikTok Technology Ltd. (organized under the laws of Ireland), TikTok Technology Canada Inc. (organized under the laws of British Columbia, Canada), TikTok Australia Pty. Ltd. (organized under the laws of New South Wales, Australia) and TikTok Hong Kong Ltd. (organized under the laws of Hong Kong). Bytedance Ltd. is organized under the laws of the Cayman Islands.

e. For each "parent, subsidiary, or other affiliate," the laws under which they currently operate, if different from the laws under which they were incorporated. If there are multiple jurisdictions, please list all that apply for each "parent, subsidiary, or other affiliate."

The TikTok platform is available in over 170 countries and regions around the world, but not mainland China. Each TikTok entity is subject to the laws applicable to the entity in their respective jurisdiction.

f. For each "parent, subsidiary, or other affiliate," the full names of the leadership of each "parent, subsidiary, or other affiliate," including the members of the board where applicable.

The Chief Executive Officer of TikTok is Shou Chew, who is based in Singapore. Other members of TikTok's leadership team include Adam Presser (head of operations and trust and safety, based in Los Angeles), Blake Chandlee (in charge of monetization business, based in Austin), and Zenia Mucha (in charge of brand and communications, based in New York). The board members of TikTok Inc. are Shou Chew and Adam Presser.

ByteDance Ltd., the ultimate parent entity of TikTok Inc. that is incorporated in the Cayman Islands, has a global board comprised of Bill Ford of General Atlantic, Arthur Dantchik of Susquehanna International Group, Philippe Laffont of Coatue, Neil Shen of HongShan (formerly Sequoia China), and the company's CEO Rubo Liang, who is based in Singapore.

g. For each "parent, subsidiary, or other affiliate," the mix of capital backing the entity, including all state-owned banks or financing regimes, or state-backed banks or financing regimes, the names of those state-owned or state-backed banks and financing regimes, and the names of those nations.

TikTok Ltd. is wholly owned by ByteDance Ltd., a Cayman Islands company. ByteDance Ltd. is majority owned by investors around the world, and the rest are owned by the founding team and employees around the world.

37. In TikTok's privacy policy is a paragraph titled "For Legal Reasons." The relevant text has been reproduced below:

How we share your information

For Legal Reasons

We may disclose any of the information we collect to respond to subpoenas, court orders, legal process, law enforcement requests, legal claims, or government inquiries, and to protect and defend the rights, interests, safety, and security of the Platform, our affiliates, users, or the public. We may also share any of the information we collect to enforce any terms applicable to the Platform, to exercise or defend any legal claims, and comply with any applicable law.

Please answer whether TikTok would consider demands made under the following to fall under the umbrella of "subpoenas, court orders, legal process, law enforcement requests, legal claims, or government inquiries" for which TikTok may "disclose any of the information [it] collect[s]":

a. China's 2014 counter-espionage law, which allows Chinese authorities to seal or seize any property linked to activities deemed harmful to the country.

No.

b. China's 2015 national security law, which outlaws threats to China's government, sovereignty and national unity as well as its economy, society, and cyber and space interests.

No.

c. China's 2017 national intelligence law, which obliges individuals, organizations, and institutions to assist Public Security and State Security officials in carrying out a wide array of "intelligence" work, and stipulates that "any organization or citizen shall support, assist, and cooperate with state intelligence work according to law."

No.

38. Of the ByteDance employees in China who have access or have previously accessed TikTok data, how many of them are affiliated with, or have some kind of relationship with, the Chinese Communist Party? Please list, in detail, what those relationships are and what data these employees had or have access to.

TikTok does not collect political affiliation information regarding its employees.

Letter Regarding the Treatment of Content from the Israel-Hamas War (October 20, 2023)

39. Describe how automated content moderation on your platform(s) has affected content from the Israel-Hamas War since October 7, 2023.

TikTok removes violative content proactively using both automated and manual processes, as well as removing violative content reported by users. Videos uploaded to TikTok are initially reviewed by TikTok's automated moderation technology, which aims to identify content that violates the Community Guidelines before it is distributed across the TikTok platform and displayed to users. These automated technology systems look at a variety of signals across content, including keywords, images, titles, descriptions, and audio, and continuously learn and adapt based on the data in each video and the moderation decisions that TikTok's human moderators ultimately make based on TikTok's Community Guidelines and related policies. If TikTok's machine-based filters identify a potential violation, the automated moderation system will either pass it on to TikTok's safety teams for further review or remove it automatically. Automatic removal is applied if there is a high degree of confidence that the content violates the Guidelines and where violations are most clear-cut. Automatic removals are subject to the user's ability to appeal that determination.

When TikTok's automated moderation systems identify potentially problematic content but cannot make an automated decision to remove it, they send the content to TikTok's safety teams for further review. To support this work, TikTok has developed technology that can identify risky or suspicious items — for example, weapons — in video frames, so that content moderators can carefully review the video and the context in which it appears. This technology improves the efficiency of TikTok's moderators by helping them more adeptly identify violative images or objects, quickly recognize violations, and make decisions accordingly.

In response to the Israel-Hamas War, TikTok has evolved its proactive automated detection systems in real-time as it identifies new threats, enabling TikTok to automatically detect and remove graphic and violent content so that neither moderators nor TikTok community members are exposed to it. TikTok identifies new emerging trends from a number of different specialized teams and sources, including: internal information points on violative content, external risks and escalations, third-party intelligence sources, and proactive efforts by its investigations teams to monitor for and identify violations.

In keeping with its commitment to ensuring fairness, TikTok notifies community members if and why their content was removed. The community member can then appeal the decision if he or she believes their content was erroneously removed, and TikTok allows community members to submit specific feedback on why they disagree with the decision to remove the content.

40. How many pieces of content from the Israel-Hamas War have been removed automatically by your systems (i.e., without any human review)?

From October 7, 2023 through December 31, 2023, TikTok automatically removed 1,234,859 videos in the conflict region for violating its Community Guidelines, including content promoting Hamas or containing hate speech, terrorism, or dangerous misinformation.

a. How many of the removals described in the previous Question were appealed?

Of the 1,234,859 videos automatically removed in the conflict region between October 7, 2023 through December 31, 2023, 147,241 video removals were appealed.

b. How many of the appeals described in the previous Question have been reviewed?

All 147,241 appeals were reviewed.

c. How many of the appeals described in the previous Question have been granted?

Review of the 147,241 appeals resulted in 80,647 videos being restored.

d. For the content described in the previous Question, do you plan to conduct a policy review of the content to ensure that content in the public interest was not erroneously removed from your platform(s)?

We recognize that some content that would otherwise violate our rules can be in the public interest to view. Public interest refers to topics that inform, inspire, or educate the community and enhance deliberation about matters of broad collective significance. We may allow content to remain on the platform one of the following public interest exceptions, such as documentary or educational. An important factor TikTok considers in looking at public interest exceptions is context, such as captions, voice over, and similar signals. TikTok encourages creators to clearly show the context to help in the review process. TikTok does not provide public interest exceptions for content that may cause extreme harm, such as showing a suicide or sexual abuse of a young person.

TikTok added opt-in screens over content relating to the Israel-Hamas War that could be shocking or graphic to help prevent people from unexpectedly viewing it as TikTok continues to make public interest exceptions for some content.

We remain agile in considering and implementing changes to both our policies and enforcement strategies. A key part of this is working with external experts, for example engaging with dozens of organizations representing Jewish and Muslim communities to help ensure our actions against antisemitism and Islamophobia are effective.

41. How would your decisions to remove content pursuant to international laws differ if you faced a legal obligation in the United States to not remove content protected by the First Amendment?

TikTok removes content that violates its Community Guidelines or applicable local law. TikTok also complies with lawful legal process. TikTok does not, however, remove non-violative content based on ideological or political considerations.

Regarding your hypothetical scenario, TikTok Inc. is a U.S. company incorporated in the United States and subject to the laws of the United States, and in the event current U.S. law regarding content moderation changes, TikTok's position is that it would comply.

Letter Regarding TikTok's Recommendation Systems (February 13, 2023)

42. Provide a complete list of the names of any individuals outside of your organization that you consulted with in developing any of the documents and information [that describe your recommendation systems and any content moderation policies for such systems].

TikTok consults with numerous experts and organizations to inform our recommendation systems and content moderation policies. We work with our Regional Safety Advisory Councils and U.S. Content Advisory Council to bring together groups of independent experts who help us develop forward-looking policies and processes that not only address the challenges of today, but

also plan ahead for the next set of issues that our industry will face. These councils are an important way to bring outside perspectives into our company and onto our platform.

Our council members represent a diverse array of backgrounds and perspectives, and are made up of experts in youth safety, free expression, hate speech, and other safety topics. They work collaboratively with us to inform and strengthen our policies, product features, and safety processes.
We have established eight regional Safety Advisory Councils in Asia Pacific, Brazil, Europe, Latin America, MENAT (Middle East, North Africa, Turkey), and a U.S. Content Advisory Council. We aim to continue expanding our regional presence. Focusing on regions allows us to better understand local challenges and trends and develop informed solutions that consider unique local context and cultures. This commitment to localization enables us to create a more targeted, responsive approach to safety and enables us to stay up-to-date with the latest developments in each region.

In addition, we regularly work with experts in online security, wellness, digital literacy, and family safety to help provide advice and resources for our community. Note that some of our partners have asked that their collaborations with TikTok remain non-public.

43. For the recommendations [viewed by users under 18], please list the top 100 sources of recommendations.

It is unclear what is meant by "top 100 sources of recommendations." The main source for the recommendation system is the user, and the primary signals for the recommendation system are user interactions. They include videos users like or share, accounts they follow, comments they post, and content they create. Other signals that the recommendation system takes into account include account and device information, such as country and language setting, and video information, which might include captions, sounds, and hashtags.

44. Please provide copies of any curation guidelines for human-curated content, accounts, or entities.

TikTok may promote or "heat" specific content (including, e.g., promoting the video of an artist who will be hosting a concert on TikTok Live) in line with company content policies to support the inclusion of diverse and high-quality content on the platform. A content operations team will review heating requests submitted by a limited number of cross-functional partners with access to the heating request, and the Content Operations team will either approve or reject the request based on their assessment of whether it follows the platform's best practices in support of content diversity and quality (including, e.g., being engaging and meaningful and focusing on timely/relevant content) and business objectives. A USDS Trust and Safety team also reviews Heating requests to help ensure no content violating its policies will be promoted using the Heating function. Even if the request is approved, increasing visibility or video views ("VV") is not guaranteed as the recommendation system will not recommend low quality content (e.g, reposted or irrelevant content). Heating impacts less than 1% of VV in the U.S. We have guidelines for the process of human-curation of content known as "heating", but do not publicly

disclose those guidelines in order to prevent bad actors from attempting to manipulate our systems.

45. Please list all U.S.-based users with more than 500,000 total followers or subscribers that have been removed from recommendations, even if temporarily, for a period of at least three continuous days within the past ten years. Please include the duration of and reason for the removal, and note whether the removal is currently in effect.

There are times we limit accounts and content for the safety of our platform and to prevent harm. Video views will vary from video to video, and the number of followers a user has or whether they have other high-performing videos does not affect how a video is suggested in For You feeds. This means that a user may see some of their videos receive more views than others. Recommendations are not determined based on how many followers a user has.

TikTok uses the same content moderation practices for Government, Politician, and Political Party Accounts and news entities as we do for other TikTok accounts. This means we'll remove any violative content and permanently remove the account for any single severe content violation, such as showing real-world violence or torture. However, because of the role these public interest accounts play in civic processes and civil society, we enforce different account restrictions in keeping with our commitment to human rights and free expression. If they reach the strike limit set for all accounts, they'll be temporarily ineligible to appear in the For You and Following feeds for 90 days.

In some cases, where public interest accounts may present a particularly high risk to public safety—such as during periods of civil unrest, elections, or other high-risk social and political environments—we may impose other restrictions. If a public interest account posts content during high-risk times that promotes violence, hate, or misinformation that could undermine a civic process or contribute to real-world harm, we may restrict that account from posting content for a period of 7 to 30 days, depending on the severity of the violation and surrounding risk. We may extend the restriction period if we determine that the actions of the account owner indicate they're likely to continue violations and public safety is still at high risk. We may also consider behavior outside of TikTok in our decision.

46. What percentage of U.S.-based recommendations on your platform(s) are political in nature, such as accounts of political figures or content discussing current political issues? If you do not include political content in recommendations, please (a) elaborate on why not and (b) provide your precision rate for enforcing this rule.

TikTok does not have a "political" content label and is therefore unable to assess what percentage of U.S.-based recommendations are political in nature. Content that is political in nature is able to be recommended to users. Like all content, it is subject to the Community Guidelines.

47. Please list the top 100 sources of political content shown in recommendations, as defined by total distribution from recommendations, for each year over the past ten years. Please provide these lists regardless of whether you have a policy to not include political content in recommendations.

In general, TikTok's labeling efforts are used to ensure that 1) content that is potentially harmful to younger users is restricted from their For You feed, 2) low visual quality content is not recommended, and 3) the For You feed is diverse. TikTok does not label content as "political." The definition of "political" is subjective, and therefore the Company does not label incoming or uploaded content as "political" in the way it may label "games" or "animals" or "travel" and other visual activities that can be seen in a particular video. Because this is not tracked, TikTok is unable to generate a list of the top 100 sources.

TikTok's Responses to Questions for the Record from Senator Hawley
Submitted March 7, 2024

1. **Do you allow your children to use social media? If so, please explain under what conditions.**

TikTok encourages each family to make their own decisions about when their children are ready to engage online, and provides a detailed Guardians' Guide[1] to help parents discuss these topics with their teens.

2. **Do you believe that children under the age of 18 should be allowed to use social media?**

Yes, we believe people under age of 18 have a right to freedom of speech and expression, and in general, we believe that parents and guardians, not government, should decide to what extent their children should be allowed to use social media.

3. **How many individuals does your company employ in Trust & Safety?**

We currently have more than 40,000 trust and safety professionals working to protect our community.

4. **How many individuals does your company employ to review content for so-called "misinformation," "disinformation," or "malinformation"?**

TikTok's community is protected by over 40,000 global trust and safety professionals. We deploy a combination of automated technology and skilled human moderators who take into account additional context and nuance which may not always be picked up by technology. All of our moderators are trained to identify misinformation and disinformation content, and we have thousands of moderators focused on moderating U.S. content including in places like Nashville, Tennessee and Austin, Texas. Some violations require further work by specialized moderators in specialized queues.

5. **How many dollars per year does your company spend on salaries for Trust & Safety officers?**

As a privately held company, TikTok does not disclose revenue or detailed budget breakdowns.

[1] https://www.tiktok.com/safety/en/guardians-guide/

6. **Do you believe that the algorithms your company has developed to sort users' feeds are protected by Section 230 of the Communications Decency Act of 1995? If so, please explain why.**

Section 230 of the Communications Decency Act ("CDA"), immunizes internet-based services from claims that treat them as a "publisher or speaker" of "information" provided by a third party. The plain meaning of "publishing" includes selecting and organizing information for display to make that information useful for an audience. Section 230 does not distinguish between publishing functions that use an algorithm and those that do not. Algorithms are everywhere, used by nearly all (if not all) search engines and other websites, and are indispensable tools for organizing the vast quantities of data available on the internet and publishing that data in a user-friendly format. Without them, the modern internet could not function.

7. **Do you believe that the algorithms your company has developed to sort users' feeds are expressive speech protected by the First Amendment to the U.S. Constitution? If so, please explain why.**

The First Amendment protects those who disseminate speech created by others.

8. **Is your company a member of a party, an amicus, or a member of an amicus in *NetChoice, LLC v. Paxton,* No. 22-555 (U.S.), or did your company provide any funds or donations to any party or amicus in that case? If so, please describe the amount of funds or donations made and the context.**

Yes, TikTok is a member of NetChoice and financially supports its work. TikTok also has provided financial support to various amici who have appeared in the case, though TikTok has not provided any amicus with funding specifically to be used in support of this litigation.

9. **Do you believe that the First Amendment to the U.S. Constitution precludes Congress from enacting legislation holding social media companies liable to users for torts they commit?**

The First Amendment to the United States Constitution provides that "Congress shall make no law . . . abridging the freedom of speech, or of the press." This language restricts the government's ability to constrain the speech of citizens. But it does not apply to conduct that does not constitute speech or expressive conduct, and even as to speech and expressive conduct, its prohibitions are not absolute.

10. **Do you believe that companies can be trusted to develop artificial general intelligence (AI) through open-source methods?**

We believe open-source does have an important place in AI development. Companies have been able to use open-source methods to make significant and trustworthy software and technology developments across multiple sectors for decades.

11. Do you believe the government should play a role in licensing certain artificial intelligence technologies, such as generative AI products?

We support government efforts to make laws and regulate artificial intelligence technologies, including generative AI products.

12. Do you believe that artificial intelligence represents an existential threat to humanity?

No. Artificial intelligence has tremendous potential to increase creativity and productivity while also presenting challenges that should be identified and mitigated.

13. Do you think that the development of large language models by Microsoft, Google, Meta, and other large companies raises antitrust concerns?

We cannot say whether the conduct of such companies raises antitrust concerns; we leave such determinations up to appropriate legal authorities.

14. What steps does your company take to make transparent the algorithms by which users are censored, shadow banned, or demonetized?

We disagree with the characterization of our moderation efforts. TikTok provides users with information on how[2] our recommendation system works and offers users the ability to learn why[3] a specific video was recommended. We ban accounts and videos in accordance with our Community Guidelines[4], Terms of Service[5] and other company policies. Video views will vary from video to video, and the number of followers or whether a creator has other high-performing videos does not affect how a video is suggested in For You feeds.

We do take measures to keep our users and platform safe, like enforcing eligibility standards for content recommended to the For You feed (which we share as part of our Community Guidelines). There are times we limit accounts and content for the safety of our platform and to prevent harm. Video views will vary from video to video, and the number of followers a user has or whether they have other high-performing videos does not affect how a video is suggested in For You feeds. This means that a user may see some of their videos receive more views than others. Recommendations are not determined based on how many followers a user has.

Users are also given the option to review TikTok's reasoning and appeal the decision if TikTok determines a video isn't eligible for the For You feed. A user must have analytics in their TikTok settings turned on to view TikTok's decision.

---

[2] https://www.tiktok.com/transparency/en-us/recommendation-system/
[3] https://newsroom.tiktok.com/en-us/learn-why-a-video-is-recommended-for-you
[4] https://www.tiktok.com/community-guidelines/en/
[5] https://www.tiktok.com/legal/page/us/terms-of-service/en

15. What steps does your company take to ensure that your company is not disproportionately targeting or censoring conservative voices?

TikTok is committed to being a platform for free expression and not being manipulated by any government. The content people see on TikTok is generated by our community and subject to our publicly available Community Guidelines and Advertising Policies, and recommendations are based on the content people have previously engaged with.

To help ensure a safe, trustworthy, and vibrant experience, TikTok maintains a set of Community Guidelines that include rules and standards for using TikTok. The guidelines apply to everyone and everything on our platform.

In keeping with our commitment to ensuring procedural fairness, TikTok notifies community members if and why their content was removed. The community member can then appeal the decision if he or she believes their content was erroneously removed, and TikTok allows community members to submit specific feedback on why they disagree with the decision to remove the content.

16. Do you condemn Hamas' terrorist attacks on the State of Israel on October 7, 2023?

TikTok stands against terrorism and is deeply saddened by the events that took place on October 7th and the ongoing Israel-Hamas War. TikTok's priority is to keep both its global community on the platform and those impacted by these tragic events safe. While it is important to stay informed about the ongoing crisis, we want to be clear that we do not allow violent, hateful, or misleading content on the platform.

TikTok has strict rules prohibiting violent threats and incitement to violence, including rules prohibiting the presence of extremist and hateful organizations on the platform. TikTok does not allow Hamas, Hezbollah, or Palestinian Islamic Jihad an organizational presence on the platform (e.g., the organizations and their prominent leaders are prohibited from holding accounts). Furthermore, TikTok prohibits the praising or glorification of Al-qassam brigade (Hamas' military wing), the Hezbollah military, and Palestinian Islamic Jihad.

From the start of the war through to the end of last year, we have removed more than 1.5 million videos and suspended more than 46,000 livestreams in Israel and Palestine for violating our Community Guidelines, including content promoting Hamas, hate speech, terrorism and misinformation.

17. What role do you believe social media companies have in promoting or limiting public speech regarding the events of October 7, 2023?

As mentioned, it is our priority to keep our global community and those impacted by the Israel-Hamas war safe. We are committed to transparency, and we remain focused on supporting free expression, upholding our commitment to human rights, and protecting our platform during the Israel-Hamas war.

To help ensure a safe, trustworthy, and vibrant experience, we maintain a set of Community Guidelines[6] that include rules and standards for using TikTok. The guidelines apply to everyone and everything on our platform. They are informed by international legal frameworks, industry best practices, and input from our community, safety and public health experts, and our regional Advisory Councils. We evolve them to address emerging risks and potential harms that may occur from new behaviors.

We also have eight guiding Community Principles[7] that help embody our commitment to human rights[8]. Our principles are centered on balancing expression with harm prevention, embracing human dignity, and ensuring our actions are fair. They shape our day-to-day work and guide how we approach difficult enforcement decisions.

Following the events that took place on October 7th, we immediately mobilized significant resources and personnel to help maintain the safety of our community and integrity of our platform. We quickly launched a command center and stepped up additional dedicated resources from within our global teams to help prevent violent, hateful, or misleading content from spreading. We do not allow any hateful behavior, hate speech, or promotion of hateful ideologies on the platform, including antisemitism and Islamophobia. We also implemented multiple proactive safety strategies to quickly identify and take action on content or accounts that violate TikTok's Community Guidelines.

In addition to launching our command center, TikTok has taken the following actions to respond to this crisis:

- Evolving our proactive automated detection systems in real-time as we identify new threats; this enables TikTok to automatically detect and remove graphic and violent content so that neither moderators nor TikTok community members are exposed to it.
- Adding more moderators who speak Arabic and Hebrew to review content related to these events. As we continue to focus on moderator care, TikTok is deploying additional well-being resources for frontline moderators through this time.
- Continuing to enforce our policies against violence, hate, and harmful misinformation by taking action to remove violative content and accounts. For example, consistent with our policies, we remove content that supports the attacks or mocks victims affected by the violence. If we identify content that is posted depicting a person who has been taken hostage, we will do everything we can to protect that individual's dignity and remove content that violates our Community Guidelines. TikTok does not tolerate attempts to incite violence or spread hateful ideologies. Our policies prohibit content praising violent and hateful organizations and individuals, and those organizations and individuals are not allowed on the platform. TikTok also blocks hashtags that promote violence or otherwise violate our Community Guidelines.

[6] https://www.tiktok.com/community-guidelines/en/overview/
[7] https://www.tiktok.com/community-guidelines/en/community-principles/
[8] https://www.tiktok.com/transparency/upholding-human-rights/

- Adding opt-in screens over content that could be shocking or graphic and which falls within public interest exceptions, to help prevent unexpected viewing of such content. We recognize that some content that may otherwise violate our Community Guidelines can be in the public interest, and we allow this content to remain on the platform for documentary, educational, and counterspeech purposes.
- Making temporary adjustments to policies that govern TikTok features in an effort to proactively prevent them from being used for hateful or violent behavior in the region. For example, TikTok added additional restrictions on the ability to livestream as a temporary measure given the heightened safety risk in the context of the current hostage situation.
- Cooperating with law enforcement agencies globally consistent with TikTok's Law Enforcement Guidelines, which are informed by legal and human rights standards. We are aware of the specific and imminent risks to human life involved in the kidnapping of hostages and work with law enforcement to ensure the safety of the victims in accordance with emergency procedures.
- Engaging with experts across industry and civil society, such as Tech Against Terrorism and our Advisory Councils, to further safeguard and secure our platform during these difficult times.
- Launching a new anti-hate and discrimination task force within our Trust and Safety team to develop an aggressive plan to further crack down on hateful behavior, with a particular focus on antisemitism and Islamophobia. As part of this effort, we invested more resources to proactively identify new and emerging trends before they gained visibility, and partnered with experts to strengthen training for moderators to address implicit bias and the unique aspects of hateful ideologies.
- Launching our #SwipeOutHate campaign in the U.S., encouraging our community to stand together against hate by reporting it in-app. The videos have already received millions of views.
- Launching Comment Care Mode, a new set of comment filters, to everyone in the U.S., Israel, and Palestine, as we continue to test the feature globally.
- Ramping up our efforts to onboard partners to our Community Partner Channel - a direct avenue for trusted flaggers around the world, including in the conflict region, to report content to us for review, which sits alongside our in-app reporting function. Since December, we have onboarded eight new organizations, including in Australia, Mexico and Denmark, representing communities affected by the war.

We continue to diligently and robustly enforce our Community Guidelines. From the start of the war through to the end of last year, we have removed more than 1.5 million videos and suspended more than 46,000 livestreams in Israel and Palestine for violating our Community Guidelines, including content promoting Hamas, hate speech, terrorism and misinformation. Globally, in the same time period, we have removed tens of millions of pieces of content and have prevented teen accounts from viewing more than 1.5 million videos containing violence or graphic content. We also remain vigilant against deceptive behaviors; from October 7 through to

the end of last year, we have removed more than 169 million fake accounts globally, and we have removed about 1.2 million bot comments on content tagged with hashtags related to the conflict.

18. What investments has your company made in anti-CSAM technology?

TikTok has a zero-tolerance approach to content that violates our youth safety policies, especially online child sexual exploitation and abuse ("CSEA") and the sharing of child sexual abuse material ("CSAM"). Any content, including animation or digitally created or manipulated media, that depicts abuse, harm, exploitation, or endangerment of minors is a violation of our Community Guidelines and will be removed when detected. To help identify CSAM, we use multiple technologies, including our own systems and hash-matching software like Microsoft's PhotoDNA, Google's Content Safety API, and YouTube's CSAI Match. With our partners, we have built hash databases – hubs of unique digital codes that have been linked to known CSAM. This means that should someone attempt to upload CSAM to TikTok that matches a unique fingerprint from a database, or if we've identified new or suspected CSAM, that content will be removed, reported to the National Center for Missing & Exploited Children ("NCMEC")[9] and to relevant law enforcement authorities. The account that shared the content will also be permanently banned.

At TikTok, we believe that collaboration is critical to solving today's most pressing challenges, including tackling CSAM. We work with leading youth safety organizations such as the Family Online Safety Institute ("FOSI")[10], Technology Coalition[11], NCMEC[12], Internet Watch Foundation ("IWF")[13], WePROTECT Global Alliance[14], and dedicated child safeguarding units in national and international law enforcement agencies. These partnerships provide opportunities to collaborate, learn and share best practices with our peers and to make critical progress toward our shared goal of ending online child sexual abuse. More information can be found here[15]. TikTok also participates in NCMEC's Take It Down[16] service, which is a free service that can help process requests to remove or stop the online sharing of nude, partially nude, or sexually explicit images or videos taken when someone is under 18 years old.

TikTok is investing in language analysis and other technical tools to understand evolving predator behaviors across all regions, including the U.S. We are building models to detect grooming and predator behaviors across the TikTok platform, as well as feature-specific models trained to better detect unique behaviors within a particular feature (e.g., DM, Live). We expect to launch these models in H2 2024. We also continuously refresh our keywords with new terms to reflect evolving behaviors and develop new strategies to help reduce the risk of predatory interactions.

---

[9] https://www.missingkids.org/home
[10] https://www.fosi.org/
[11] https://www.technologycoalition.org/
[12] https://www.missingkids.org/home
[13] https://www.iwf.org.uk/
[14] https://www.weprotect.org/
[15] https://www.tiktok.com/transparency/en-us/protecting-teens/
[16] https://takeitdown.ncmec.org/

We also build age-appropriate experiences and controls that enable teens to have a safe space to create, share, discover, and connect. We're dedicated to partnering with families in this work as we share their interests in supporting teens as they start to explore the online world independently.

19. Have you read the Fifth Circuit's opinion in *Missouri v. Biden*, No. 23-30445?

TikTok is familiar with the Fifth Circuit's opinion in *Missouri v. Biden*, No. 23-30445.

20. Do you dispute any factual findings in the Fifth Circuit's opinions or the district court's opinions?

The Fifth Circuit's opinion in *Missouri v. Biden*, No. 23-30445, does not make any factual findings regarding TikTok. The district court's opinion notes that TikTok participated in the "Election Integrity Partnership" ahead of the November 2020 election, *Missouri v. Biden*, No. 3:22-CV-01213 (W.D. La. July 4, 2023), which is accurate.

21. Does your platform continue to receive requests from federal agencies to censor or promote certain content?

TikTok is committed to being a platform for free expression and not being manipulated by any government. The content people see on TikTok is generated by our community and subject to our publicly available Community Guidelines and Advertising Policies, and recommendations are based on the content people have previously engaged with.

TikTok does provide formal processes for governmental requests to remove or restrict content from the platform. We honor valid law enforcement requests through the proper channels and where otherwise required by law. When TikTok receives such requests from government authorities, we review and take action upon content in line with our Community Guidelines, Terms of Service, and applicable law. However, TikTok does not remove or limit the distribution of content solely because someone—whether it be a user, a member of the general public, or a government official—asks. If we believe the content at issue in a request doesn't violate the standard laid out in our Community Guidelines but violates applicable law, we may restrict the availability of the reported content in the country where it is considered to be illegal. If we believe that a request isn't legally valid or does not violate our Community Guidelines, Terms of Service, or applicable law, we will reject the request. Twice a year, we disclose such requests, as well as actions taken in response to those requests, in our Government Removal Requests Reports[17].

Separate and apart from content moderation, TikTok also responds to law enforcement requests, including requests for disclosure or preservation of user information. We evaluate these requests pursuant to our publicly available Law Enforcement Guidelines[18] and provide information in

[17] https://www.tiktok.com/transparency/en-us/government-removal-requests-2023-1/
[18] https://www.tiktok.com/legal/page/global/law-enforcement/en

response to valid legal process or in emergency circumstances. Twice a year, we publish information about law enforcement requests, as well as actions taken in response to those requests, in our Information Requests Reports[19].

22. What steps do your platforms take to verify and enforce age restrictions?

TikTok's goal of providing an age-appropriate experience to its users begins with an industry-standard neutral age gate that is consistent with the Federal Trade Commission's ("FTC") guidance for age verification under the Children's Online Privacy Protection Act ("COPPA"). If an individual selects a birthdate that indicates that they are under the age of 13 when creating a TikTok account in the U.S., they are directed to TikTok's under 13 experience, where they can watch a curated library of age-appropriate videos. In addition to being restricted to only certain approved content, users in the under 13 experience cannot access many of the features and functions that are available to users on the 13+ experience. For example, they are not able to post videos on the platform, comment on videos, message other users, maintain a profile or followers, receive ads, or be directed off the TikTok platform.

Beyond age gate, TikTok uses technologies and human moderators, as well as user and third party reporting, to detect and remove users in the 13+ experience who are suspected to be under 13. An account in the 13+ experience that is flagged as being potentially under 13 is routed to a dedicated team of trained moderators who would review the account to determine if it should be banned for not meeting the minimum age requirement. If the moderator makes a determination that the account belongs to a suspected underage user, the account would be removed from the 13+ experience.

23. In cases where a child's safety is at risk, how does your company collaborate with law enforcement? What information or assistance is provided?

If a minor's safety is at risk, law enforcement receives guidance through our Law Enforcement Guidelines[20] and we have provided training on how to utilize our emergency request process and form[21]. After submitting this form, requests are received by our 24/7 on-call Emergency Response Team specialists who evaluate the request. If the request meets our standards of being imminent, specific and credible, TikTok will produce user data to help locate or identify the individuals at issue. For example, during our presentation at the National Law Enforcement Training on Child Exploitation conference, our Law Enforcement Outreach team was approached by members of the Georgia Bureau of Investigations ("GBI") regarding a pressing issue involving Child Sexual Abuse Material ("CSAM"). Recognizing the urgency of the situation, our team swiftly mobilized our Emergency Response Team ("ERT") and Child Safety Team to provide assistance. Through collaborative efforts and the seamless coordination of resources, TikTok was able to furnish the GBI with crucial information essential for their investigation. This timely support facilitated swift action by the GBI and the Atlanta Police

[19] https://www.tiktok.com/transparency/en-us/information-requests-2023-1/
[20] https://www.tiktok.com/legal/page/global/law-enforcement/en
[21] https://law-enforcement-tiktok-uscan.zendesk.com/hc/en-us/requests/new?ticket_form_id=4416470231707

Department ("APD"), leading to the successful location and recovery of the child within a remarkably short timeframe, just one day after our involvement.
Outside of emergencies, TikTok proactively engages with law enforcement through its Law Enforcement Outreach team, which provides training, assistance, and points of contact to law enforcement authorities across the country. In the last year, our LE Outreach team provided such training and support to the numerous child safety focused organizations, task forces, and law enforcement agencies, including ICAC Commanders, HSI NYO Trafficking and Child Exploitation Task Force, DOJ SMART Unit (Sex Offenders Monitoring, Apprehending, Registering, and Tracking), as well as numerous engagements at conferences and at the state level.

Upon identification of imminent risk (i.e., sextortion or on-going abuse cases), TikTok's Child Safety Team completes a CyberTip Report submission to NCMEC in by utilizing their ESP escalation function to prioritize the report. Additionally, TikTok's Law Enforcement Outreach Team is also made aware of the referral, allowing them to interface with local law enforcement directly. If the user is within the United States, the Emergency Response team is also notified and will review and assess the request immediately. It is common practice that ERT will review and assess the request within the minute of receiving the request.

24. Do you believe there is any expressive value in CGI or AI generated CSAM?

CGI or AI-generated CSAM is abhorrent and TikTok condemns such content and those individuals who generate such content in the strongest possible terms. Such content violates TikTok's Community Guidelines. TikTok proactively and aggressively works to prevent such content from appearing on its platform and to quickly remove any such content that appears on its platform notwithstanding the significant efforts and substantial resources it devotes to moderating such content. TikTok has also developed features to enable the reporting of potential CSAM.

25. Do you believe that CGI or AI generated CSAM is protected by the First Amendment to the U.S. Constitution?

CSAM is not protected by the First Amendment. All CSAM is abhorrent and TikTok condemns such content and those individuals who generate such content in the strongest possible terms. Such content violates TikTok's Community Guidelines. TikTok proactively and aggressively works to prevent such content from appearing on its platform and to quickly remove any such content found on its platform. TikTok has also developed features to enable the in-app reporting of potential CSAM.

26. What measures does your platform take to ensure that children only see age-appropriate advertisements?

TikTok has policies in place on advertisements for users under the age of 18.

27. Will you commit to setting up a compensation fund for those who have been harmed by your platform?

TikTok disagrees with the assertion that its platform is harmful to users. TikTok evaluates user complaints on a case-by-case basis and will continue to do so, consistent with its commitment to partnering with our community to help ensure people have a safe and positive experience on the platform.

28. Do you believe that the First Amendment to the U.S. Constitution precludes Congress from enacting a law to ban TikTok on U.S. devices?

Yes.

29. Buzzfeed and Forbes have repeatedly reported instances where ByteDance has accessed the data of U.S. citizens. Do you dispute the veracity of these reports?

As TikTok has stated publicly, TikTok has historically stored U.S. user data collected by the TikTok app in its own data centers in the U.S. and Singapore. These data centers in the U.S. and Singapore have served as the default storage location for the core databases that support the TikTok platform.

TikTok has also been clear that as a global company with a global workforce, access to U.S. user data has historically been made available to employees based on their job function and demonstrated need to perform their roles. In 2020, TikTok set a goal to limit data access across regions, and as detailed below, has made significant progress on this commitment. TikTok has been working earnestly since the launch of Project Texas on a phased initiative to strengthen TikTok's data protection policies and protocols, further protect our users, and build confidence in our systems and controls in the United States. This includes the creation of a standalone entity called TikTok U.S. Data Security Inc. ("USDS"). USDS is a special purpose subsidiary tasked with managing all business functions that require access to "protected U.S. user data" and safeguarding the systems that deliver the TikTok app experience in the U.S. Our approach under this initiative is to set controls to block transfer of "protected U.S. user data" outside of the U.S.-based Oracle and USDS Infrastructure, nor will it be accessible by non-USDS employees, with limited exceptions explained in more detail below.

Since January of 2023, new protected U.S. user data has been stored in the Oracle Cloud in an environment controlled by USDS. We've taken steps to limit access to protected U.S. user data in the Oracle environment, unless authorization is given by USDS pursuant to limited exceptions, such as for legal and compliance purposes. Traffic from the Oracle Cloud now goes through Oracle controlled gateways to prevent protected U.S. user data from being transferred to or accessed by employees of TikTok or ByteDance.

In March 2023, TikTok began the process of deleting historic protected U.S. user data globally; once that process is complete, we will have completed our process to end all access to protected U.S. user data outside of USDS except under limited circumstances stipulated in our draft

national security agreement, subject to any further changes prior to finalization of such agreement.

TikTok has been clear that there are certain, limited exceptions to the definition of protected data. These exceptions are in place to help ensure interoperability of TikTok as a global platform and were determined as part of TikTok's extensive, multi-year discussions with CFIUS that have occurred under two Administrations. Exceptions include categories such as business metrics, interoperability data, and certain creator data, if a creator voluntarily signs up for a commercial program to be supported by TikTok in reaching new audiences and monetizing content. As part of Project Texas, we have also designed and operationalized a controlled mechanism to enable TikTok to respond to global government and litigation demands for documents relating to users.

30. Forbes has reported that TikTok has used user data to surveil U.S. journalists. Do you dispute the veracity of these reports?

As the company has disclosed previously, in late 2022 the company learned that certain employees had accessed TikTok user data in an unsuccessful and misguided attempt to trace the source of leaks of confidential TikTok information. The company has seen no evidence that live, real-time user data or precise location data was accessed as part of these efforts.

31. Do you condemn the actions by the Chinese Communist Party in perpetrating a genocide of the Uyghur people?

TikTok provides a global entertainment app experience and is not a political organization. We are committed to respecting the human rights of all people impacted by our platform, regardless of their ethnicity, orientation, background, or life experience. As a global entertainment company, TikTok has a responsibility to ensure that our community is treated with fundamental dignity and respect on our platform. A wide variety of content can be found on TikTok, including videos criticizing the actions by the Chinese government against the Uyghur people.

32. TikTok is wholly owned by ByteDance, correct?

The ultimate parent company of TikTok Inc. is ByteDance Ltd., a privately-owned holding company established in the Cayman Islands. ByteDance Ltd. is majority owned by investors around the world, and the rest of the shares are owned by the founding team and employees around the world. ByteDance Ltd.'s Board of Directors is comprised of five individuals, three of whom are American.

33. Project Texas is undertaken at the direction of ByteDance?

The TikTok operating companies are subsidiaries of ByteDance Ltd. In response to national security concerns raised by CFIUS, TikTok launched an initiative to build a secure environment for protected U.S. user data, to ensure the platform remains free from outside influence, and to implement additional safeguards on our content recommendation and moderation tools.

34. ByteDance can hire and fire TikTok directors?

As the ultimate parent company of TikTok Inc., ByteDance Ltd. has the ability to appoint and remove directors of TikTok Inc.

35. If ByteDance asked a TikTok director to divulge American user data and that director refused, ByteDance could remove them?

First, such a request would violate company policies.

Furthermore, TikTok has imposed data access policies to help ensure that adequate safeguards are in place to protect personal information. If a person has no business need to access protected U.S. user data, they are not afforded such access under the terms of those policies. In addition to these safeguards, TikTok Inc.'s U.S. subsidiary, USDS, is undertaking efforts that are unprecedented among our peer group to build a secure environment for protected U.S. user data, which include protections against the types of transfers contemplated by the question.

36. If ByteDance ordered a TikTok director to conceal ByteDance's ability to access TikTok data and that director refused, ByteDance could remove them?

As we have explained, ByteDance employees do not have access to TikTok protected U.S. user data. Similar to the previous question, such would violate company policies and be unethical.

TikTok has imposed data access policies to help ensure that adequate safeguards are in place to protect personal information. If a person has no business need to access protected U.S. user data, they are not afforded such access under the terms of those policies. In addition to these safeguards, TikTok Inc.'s U.S. subsidiary, USDS, is undertaking efforts that are unprecedented among our peer group to build a secure environment for protected U.S. user data.

37. Do you dispute that ByteDance is subject to China's National Security Law, which requires all organizations to support, assist, and cooperate with national intelligence efforts?

TikTok does not offer the TikTok app for download in mainland China. TikTok Inc. has not been asked by the Chinese government for U.S. user data. TikTok discloses on a regular basis in its Information Requests Reports[22] the volume and type of requests for user information received from governments and law enforcement agencies, and whether the data was disclosed or presented.

TikTok Inc.'s ultimate parent company is ByteDance Ltd., a privately-held, global holding company incorporated in the Cayman Islands and subject to the law of the Cayman Islands. ByteDance Ltd. owns many businesses, some of which operate in China. ByteDance is obligated to comply with China's laws insofar as they relate to the ByteDance business just as it is obligated to comply with the laws of the United States. Other U.S. companies doing business in

[22] https://www.tiktok.com/transparency/en-us/information-requests-2023-1/

China would likewise be required to comply with Chinese laws insofar as they are doing business in China.

TikTok does not make the app available for download in mainland China. Additionally, we've taken steps to limit to only USDS personnel access to protected U.S. user data in the Oracle environment, unless authorization is given by USDS pursuant to limited exceptions. Traffic from the Oracle Cloud now goes through Oracle controlled gateways to prevent protected U.S. TikTok user data from being transferred to or accessed by employees of TikTok or ByteDance. TikTok believes that these cutting-edge measures are unprecedented among our peer group and will provide many layers of protection for U.S. user data.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

38. What is your understanding of ByteDance's obligations under China's National Security and Intelligence Laws?

ByteDance is obligated to comply with China's laws insofar as they relate to the ByteDance business just as it is obligated to comply with the laws of the United States. Other U.S. companies doing business in China would likewise be required to comply with Chinese laws insofar as they are doing business in China.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

TikTok's Responses to Questions for the Record from Senator Tillis
Submitted March 7, 2024

1. **Twenty-one is the minimum age to purchase highly regulated adult products such as alcohol, tobacco, and nicotine. Nevertheless, there is a proliferation of user-generated content posted on social media sites featuring underage use of these products.**

   Recently, some have proposed banning these age-restricted products due in part to the user-generated content being available on your respective platforms. Surely, banning these products cannot be the answer. However, we must do more – your company must do more – to shield underage audiences from exposure to this content.

   Therefore, as the content moderator of these platforms, what policies do you have in place, and what more can you do, to prevent this type of user-generated content from reaching underage audiences? How do you respond to requests to pull this content from your sites when deemed inappropriate for underage audiences?

Youth safety and well-being is our priority. Per our Community Guidelines, we do not allow content that may put young people at risk of exploitation, or psychological, physical, or developmental harm. This includes child sexual abuse material ("CSAM"), youth abuse, bullying, dangerous activities and challenges, exposure to overtly mature themes, and consumption of alcohol, tobacco, drugs, or regulated substances. If we become aware of youth exploitation on our platform, we will ban the account, as well as any other accounts belonging to the person that we also discover.

While adults make personal choices about how they engage with alcohol, drugs, and tobacco, we recognize that there are risks connected to using these substances. Therefore, we do not allow showing or promoting recreational drug use, or the trade of alcohol, tobacco products, and drugs. We also recognize that using these substances can put young people at a heightened risk of harm, so we do not allow content showing or promoting young people possessing or consuming alcohol, tobacco products, and drugs. Content is also age-restricted from teens, and is ineligible for anyone's For You feed, if it shows adults consuming excessive amounts of alcohol or tobacco products.

Any advertisements on the platform must also adhere to our Community Guidelines as well as our Advertising Guidelines. We do not allow ad content and landing pages to display or promote tobacco, tobacco-related products such as cigars, tobacco pipes, rolling papers, or e-cigarettes, or smoking-related behavior in real life, including but not limited to alternatives that imitate the act of smoking. We also do not allow ad content and landing pages to display or promote illegal drugs, controlled drugs, prescription drug abuse, recreational drugs, drug paraphernalia, or

accessories or supplies any of such, including the use of them. In general, we do not allow ad content and landing pages to display, facilitate, or promote services or activities considered illegal in a given jurisdiction.

We provide options for our community to report content that they believe violates our Community Guidelines and Advertising Guidelines. Videos can be reported using our online form[1], or within the TikTok app by:

1. Going to the video you need to report.
2. Pressing and long holding on the video.
3. Selecting **Report,** tapping on '**Illegal activities and regulated goods'** and then submitting report under '**Drugs and controlled substances'**

Once videos are reported, we use a combination of human and machine moderation to review our content for Community Guideline violations, and take appropriate action. We will update the reporting individual on the status and progress of the report in their Inbox.

We also have resources within our Safety Center[2] to share information on treatment, support and recovery resources[3].

2. **Public reports conclude that drug cartels use social media like TikTok, META, X, Snapchat, and others to plan, organize, and communicate in real-time. These communications coincide directly with criminal activity.**

   What are your companies doing to crack down on cartel coordination? Specifically, in the recruitment of children to commit crimes or assist in the sale/distribution of illicit drugs?

TikTok's Community Guidelines prohibit the sale, trade, promotion, use and the depiction of drugs, including controlled substances, for both organic and paid content. Additionally, we do not promote content that discusses controlled substances (such as adult humor) in our For You feed to help ensure that it is an appropriate and comfortable place for all audiences. We also work to ensure TikTok does not enable activities that violate laws or regulations. We do not allow individuals or organizations on our platform who promote, engage in, or seek to recruit others for violent or illegal activities. We remove these individuals and organizations from our platform, including criminal organizations, such as cartels. We also remove content that promotes or enables criminal activities to prevent such behavior from being normalized, glorified, imitated, or facilitated.

We use a combination of image- and text-based AI to detect designated cartels (tools which we also use for drug and firearm detection). TikTok does not publicly disclose details about our strategies (such as keyword lists, images, or hashes), as these can be used by bad actors to

[1] https://www.tiktok.com/legal/report/feedback
[2] https://www.tiktok.com/safety/en-us/
[3] https://www.tiktok.com/safety/en-us/substance-support/

circumvent our safety strategies. Once organizations are identified, we work to identify and detect symbols, slogans, logos, and other indicators associated with groups such as cartel groups that we then remove from the platform. We remove any content – including video, audio, livestream, images, comments, links or other text – that violates our Community Guidelines, and we will suspend or ban accounts involved in severe or repeated violations.

TikTok has also established a subsidiary, TikTok U.S. Data Security Inc. ("USDS"), to control access to protected U.S. user data, content recommendation, and moderation systems. Within USDS we have teams which proactively investigate potential serious criminal conduct for law enforcement referrals. We have teams focused on illegal and regulated activities, which investigate issues in the U.S. market and identify solutions to any threats in collaboration with our policy, product, and operations teams. These teams rely on leads such as keywords, emojis, coded language, hashtags, audios, and examples of content/accounts.

We may also proactively report accounts to relevant legal authorities under certain circumstances, including where there are immediate risks of harm. For example, we forward suspected illicit drug activity content to law enforcement when it poses specific, credible and imminent threats to users or the community. In addition, this team routinely discloses relevant user data in response to valid legal requests from law enforcement agencies at the federal, state, and local levels. As an example, we have followed up on federal law enforcement leads and removed accounts for violating drug trade policies. We also take action on other accounts we surface as part of investigating these leads. This can and has led to removing dozens of other drug related accounts.

In addition to these strategies, we also have a variety of product-related features in place to protect our community:

- Accounts under the age of 16 are not allowed to use our direct messaging service, reducing the risk that they would be contacted for recruitment or promotion purposes.
- U.S. users can include URLs to external websites, but those URLs aren't clickable. That creates meaningful friction in spreading dis/misinformation and other material that violates our Community Guidelines.
- We also have a running list of websites that are blocked on platform, including those that lead to recruitment or promotional sites.

3. **What steps does your platform take to proactively remove, delist, and ban any posts, users, websites, and advertisements associated with the sale and distribution of fentanyl and other illicit drugs?**

As mentioned in our response to your Question 2, TikTok implements a multi-faceted approach to combat illegal drugs on our platform. Our Community Guidelines prohibit the sale, trade, promotion, use and the depiction of drugs includes controlled substances like fentanyl, for both organic and paid content. We also do not promote content that discusses controlled substances (such as adult humor) in our For You feed to help ensure that it is an appropriate and comfortable place for all audiences.

TikTok uses a combination of technology and human moderation to detect and remove violating content and accounts, including illegal drug-related content. Our systems work to detect weapons and suspicious accounts, as well as illicit activities on audio and livestreams. Our moderation system also uses advanced technology to identify and flag comments for evidence of solicitation or promotion of criminal activities including the sale, trade, promotion, use, or the depiction of drugs. This work is supported by third parties with expertise in various issue areas.

Our USDS teams referenced in Question 2 proactively investigate threats related to fentanyl as well and work across functions to identify solutions. We also regularly consult external experts to inform our policy development and stay on top of evolving issues. We have worked with organizations like LegitScript and Song for Charlie to spread awareness about the Fentanyl crisis and help us expand how we care for our community. We have also engaged with the Drug Enforcement Administration's ("DEA") to raise awareness of the dangers of fentanyl. For example, TikTok has received information and resources from the DEA including drug-related emojis and keywords, such as fentanyl precursors, which have subsequently been incorporated into TikTok's moderator training materials. TikTok has also engaged with the DEA on Operation Overdrive, an initiative to actively combat drug-related violence and overdoses.

4. **One area of growing concern is the sale and distribution of fake or counterfeit vaping devices online, particularly in connection with so-called Delta-8 THC. Counterfeit vapes, many coming from China, have unsafe and even potentially deadly chemicals. They have caused hospitalizations and death. What are your platforms doing to combat this problem?**

While adults make personal choices about how they engage with alcohol, drugs, and tobacco, we recognize that there are risks connected to using these substances. We do not allow showing or promoting recreational drug use, or the trade of alcohol, tobacco products, and drugs. We also recognize that using these substances can put young people at a heightened risk of harm. We do not allow showing or promoting young people possessing or consuming alcohol, tobacco products, and drugs. We define tobacco products to include vaping products, smokeless or combustible tobacco products, synthetic nicotine products, E-cigarettes, and other Electronic Nicotine Delivery Systems.

In terms of advertisements, we do not allow ad content and landing pages to display or promote tobacco, tobacco-related products such as cigars, tobacco pipes, rolling papers, or e-cigarettes, or smoking-related behavior in real life, including but not limited to alternatives that imitate the act of smoking. Additionally, alcohol, tobacco, and e-cigarettes are prohibited from being sold on our platform. This includes, but is not limited to, products containing or derived from alcohol, tobacco, nicotine, vaping substances, smokeless or combustible tobacco items, synthetic nicotine products, smoking equipment and accessories, E-cigarettes, and other Electronic Nicotine Delivery Systems.

5. **What are the main impediments your platform encounters in identifying all fentanyl and illicit drug advertisements posted to your platform(s) automatically? Please describe any circumstances in which you do not or cannot apply detection technologies against content transmitted on your platform(s).**

TikTok uses a combination of advanced technology and human moderation to detect and remove violating content and accounts, including illegal drug-related content. That said, bad actors may attempt to circumvent our protections and policies through the use of evolving signals and coded language, or by utilizing less explicit content and communicating through comments, user profiles, direct messages, or through the sharing of links and off-platform redirection.

We remain agile to combat illegal activity on the platform and respond to any quickly evolving situations. As mentioned in our response to Question 2 and Question 3, we have teams dedicated to investigating issues involving illegal and regulated activities in the U.S. market, and identifying solutions to any threats in collaboration with our policy, product, and operations teams. We also regularly consult external experts to inform our policy development and stay on top of evolving trends and issues (e.g., working with the DEA to identify including drug-related emojis and keywords, such as fentanyl precursors, and incorporating them into TikTok's moderator training materials).

6. **How many posts, users, websites, and advertisements have you removed, delisted, and banned per year for the sale and distribution of fentanyl and other illicit drugs? How many per year? Have you seen an increase in illicit drugs being advertised to children on your platform(s)?**

Per our Community Guidelines, TikTok does not allow content showing or promoting recreational drug use, or the trade of alcohol, tobacco products, and drugs. We also do not allow content showing or promoting young people possessing or consuming alcohol, tobacco products, and drugs. These policies apply to everything on our platform.

TikTok removes videos that violate these Community Guidelines. TikTok removed over 96 million videos in 2022 and over 50 million videos in H1 2023 for violations of the Illegal Activities and Regulated Goods policy, which includes removals for violations of the subpolicy on Drugs, Controlled Substances, Alcohol, and Tobacco. We updated our policies in March 2023, and similar data is now captured in a slightly broader policy on Regulated Goods and Commercial Activities[4]. In H2 2023 we removed over 107 million videos for violations of our updated Regulated Goods and Commercial Activities policy, which includes removals for violations of our new subpolicy on Alcohol, Tobacco, and Drugs. We aim to provide more granular data on removals under our subpolicies within future Community Guidelines Enforcement Reports[5], but historically, removals for violations of our alcohol, tobacco, and drug-related subpolicies represented 30-40% of the video removal figures provided.

[4] https://www.tiktok.com/community-guidelines/en/regulated-commercial-activities/
[5] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/

Violating TikTok's Community Guidelines can also lead to enforcement actions on accounts. We will ban accounts or users if they engage in: (1) a single severe content violation, (2) repeated content violations within a 90-day period, (3) circumvention, or (4) the operation of accounts dedicated to activity that violates our rules. Account bans for Community Guidelines violations may occur for multiple reasons, such as a video violation and multiple comment violations. TikTok removed over 18 million accounts in 2022 and over 14 million accounts in 2023 for violating our Community Guidelines, including our drug-related policies. Note that these numbers do not include accounts removed under suspicion of being under the age of 13. This data is publicly available in our Community Guidelines Enforcement Reports[6].

In terms of advertisements, TikTok does not allow ad content and landing pages to display or promote illegal drugs, controlled drugs, prescription drug abuse, recreational drugs, drug paraphernalia, or accessories or supplies any of such, including the use of them. In 2023, we rejected 10,184 advertisements, suspended 346 advertiser accounts, and removed 5,336 URLs within advertisements for violating our policies on drugs (note that not all advertisements contain URLs).

7. **Are there any other roadblocks or impediments that you face in addressing fentanyl and illicit drug advertisements on your platform(s), and working with law enforcement on such matters? If yes, what are they? If no, how many cases have been transmitted to law enforcement and DEA?**

As mentioned in our response to Question 2, we do have teams that coordinate with law enforcement to address this issue. These teams proactively investigate potential serious criminal conduct for law enforcement referrals. We may proactively report accounts to relevant legal authorities under certain circumstances, including where there are immediate risks of harm (e.g., suspected illicit drug activity when it poses specific, credible and imminent threats to users or the community). In addition, TikTok routinely discloses relevant user data in response to valid legal requests from law enforcement agencies at the federal, state, and local levels.

TikTok has also engaged with the DEA and DOJ on information sharing efforts. For example, TikTok has received information and resources from the DEA including drug-related emojis and keywords, such as fentanyl precursors, which have subsequently been incorporated into TikTok's moderator training materials. TikTok has also engaged with the DEA on Operation Overdrive, an initiative to actively combat drug-related violence and overdoses.

Additionally, TikTok has a robust law enforcement outreach team that is dedicated to meaningful engagement with law enforcement officers at all levels. Since 2021, this team has trained over 13,000 U.S. law enforcement officers on our platform and company. We maintain an open dialogue for questions and answers through this program, and provide resources and contacts for questions and emergencies.

[6] https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-3/

8. **How do you work with organizations, advocates, and experts focused on drug prevention and addiction recovery to adapt your products and operations to keep up with the illicit drug crisis — including working with parents that have lost children due to lethal drugs bought online?**

TikTok regularly consults external experts to inform our policy development and stay on top of evolving issues. We have worked with organizations like LegitScript and Song for Charlie to spread awareness about the Fentanyl crisis, including on our platform, and to help expand how we care for our community. As mentioned in Question 2, we have also engaged with the DEA to raise awareness of the dangers of fentanyl and to receive information to incorporate into our moderator training materials.

TikTok also prioritizes user education, and we recently rolled out a Substance Support Safety Center[7] to provide information about the harmful effects of substances, what to do if someone feels pressured to take substances, and direct contact information for the Substance Abuse and Mental Health Services Administration and Crisis Text Line. For example, our Substance Support page highlights the DEA's *One Pill Can Kill* campaign, which offers information on fake pills that are marketed as legitimate prescription pills and may be deadly.

9. **What are the total number of meetings that your company has had with parents to address online safety concerns? Can you provide the total number of meetings over the last three years? Please separate this last question's answer by number per year.**

While our teams around the world regularly engage in these conversations, we do not collect information about the number of such meetings in a centralized manner and are unable to provide a detailed response.

10. **In 2022, then National Center for Missing & Exploited Children (NCMEC) received over 32 million reports of Child Sexual Abuse Material (CSAM). Reports of online sex crimes to the CyberTipline are growing exponentially year by year. Out of those 32 million reports, how many did your platform submit to NCMEC?**

As represented on NCMEC's website, we submitted roughly 288,000 reports[8] in 2022, and were alone among our peers in receiving only one request from NCMEC[9] that year.

11. **There is concern that this number is going to fall dramatically this year because of the adoption of end-to-end encryption, not because the problem is going away. How will your company track and address this issue moving forward?**

TikTok has not adopted end-to-end encryption.

---

[7] https://www.tiktok.com/safety/en/substance-support/
[8] https://www.missingkids.org/content/dam/missingkids/pdfs/2022-reports-by-esp.pdf
[9] https://www.missingkids.org/content/dam/missingkids/pdfs/2022-notifications-by-ncmec-per-esp.pdf

12. Has your platform seen an increase of suspected online child sexual exploitation-CSAM over the past few years? If so, what do you believe is the driving factor on why it's happening on your platform?

TikTok has seen an increase in reporting since the platform launched, while also seeing an increase in the numbers of people joining the TikTok community. As we have matured, we have also worked to improve our detection systems and community tools.

13. What are some new tools or strategies that your platform has implemented to identify CSAM? How closely does your platform work with NCMEC?

TikTok has a zero-tolerance approach to content that violates our youth safety policies, especially online child sexual exploitation and abuse ("CSEA") and the sharing of child sexual abuse material ("CSAM"). Any content, including animation or digitally created or manipulated media, that depicts abuse, harm, exploitation, or endangerment of minors is a violation of our Community Guidelines and will be removed when detected. To help identify CSAM, we use multiple technologies, including our own systems and hash-matching software like Microsoft's PhotoDNA, Google's Content Safety API, and YouTube's CSAI Match. With our partners, we have built hash databases – hubs of unique digital codes that have been linked to known CSAM. This means that should someone attempt to upload CSAM to TikTok that matches a unique fingerprint from a database, or if we've identified new or suspected CSAM, that content will be removed, reported to the National Center for Missing & Exploited Children ("NCMEC") and to relevant law enforcement authorities. The account that shared the content will also be permanently banned.

At TikTok, we believe that collaboration is critical to solving today's most pressing challenges, including tackling CSAM. We work with leading youth safety organizations such as the Family Online Safety Institute ("FOSI")[10], Technology Coalition[11], NCMEC[12], Internet Watch Foundation ("IWF")[13], WePROTECT Global Alliance[14], and dedicated child safeguarding units in national and international law enforcement agencies. These partnerships provide opportunities to collaborate, learn and share best practices with our peers and to make critical progress toward our shared goal of ending online child sexual abuse. More information can be found here.

TikTok has a close operational relationship with NCMEC, which includes monthly meetings to calibrate trends and current issues. As a result of these conversations, an API reporting tool was developed to streamline and scale CyberTip submissions. TikTok also participates in NCMEC's Take It Down service,[15] which is a free service that can help process requests to remove or stop the online sharing of nude, partially nude, or sexually explicit images or videos taken when someone is under 18 years old.

[10] https://www.fosi.org/
[11] https://www.technologycoalition.org/
[12] https://www.missingkids.org/home
[13] https://www.iwf.org.uk/
[14] https://www.weprotect.org/
[15] https://takeitdown.ncmec.org

We also build age-appropriate experiences and controls that enable teens to have a safe space to create, share, discover, and connect. We're dedicated to partnering with families in this work as we share their interests in supporting teens as they start to explore the online world independently.

14. What resources or help does your platform provide to victims of CSAM? Does your platform work with local victim groups and professionals?

TikTok has developed a safety center for survivors of sexual abuse[16] that provides resources for help and information in more than 30 countries. The resources can be accessed either by going directly to the safety center or by search terms/keywords associated with abuse through an in-app response that redirect our community members to supportive resources. The pages also links to StopNCII.org, which allows people to report non-consensual sexual images (also referred to as image-based sexual abuse). StopNCII translates those images to hashes and shares them with TikTok and other companies to remove them from the app. TikTok also participates in NCMEC's Take It Down service.[17]

15. What are the top technical hurdles your company faces in combatting CSAM?

TikTok has a zero-tolerance approach to content that violates our youth safety policies, especially online child sexual exploitation and abuse ("CSEA") and the sharing of child sexual abuse material ("CSAM"). Any content, including animation or digitally created or manipulated media, that depicts abuse, harm, exploitation, or endangerment of minors is a violation of our Community Guidelines and will be removed when detected. To help identify CSAM, we use multiple technologies, including our own systems and hash-matching software like Microsoft's PhotoDNA, Google's Content Safety API, and YouTube's CSAI Match. With our partners, we have built hash databases – hubs of unique digital codes that have been linked to known CSAM. This means that should someone attempt to upload CSAM to TikTok that matches a unique fingerprint from a database, or if we've identified new or suspected CSAM, that content will be removed, reported to the National Center for Missing & Exploited Children ("NCMEC") and to relevant law enforcement authorities. The account that shared the content will also be permanently banned.

As a result of rapidly evolving technology in this space, including that utilized by bad actors, significant investments are needed to advance platforms' ability to detect and swiftly remove this content. Additionally, data deletion requirements can limit our available training data. Industry solutions for CSAM detection are more widely standardized for images and for previously discovered content (e.g., PhotoDNA). Establishing industry solutions for video detection and for new content detection would be highly impactful given the nature of our platform, but development of such solutions would be subject to technical challenges associated with higher complexity media types as well as training data sensitivity.

[16] https://www.tiktok.com/safety/en/sexual-assault-resources/
[17] https://takeitdown.ncmec.org

16. There seem to be competing views on how to regulate algorithms. Some suggest that more transparency is needed, while others want more privacy. Can you provide your perspective on whether more or less transparency is needed when it comes to algorithms?

There is no single answer to this question as algorithms are used by companies across industries and in myriad ways to deliver value to customers and address potential risks and harms. TikTok has a commitment towards transparency for how our algorithms work (see our explainer videos on how our recommendation system works[18] and why a specific video was recommended[19]) and also a commitment towards strengthening the security and privacy of our users[20]. We've also launched a Research API[21] that we've built as part of our efforts to stay accountable. The API provides researchers with access to public data on content and accounts on our platform in a way that preserves individual user privacy.

17. Do you believe that large companies and platforms like yours can use algorithms to stifle innovation or small businesses?

Algorithms are one of many basic technology tools that individuals or organizations of any size can use to be competitive in their respective markets.

18. What do you believe is the role of government in regulating algorithms? What, if any, unintended consequences would there be if Congress gets involved?

Algorithms are one of many basic technology tools that individuals or organizations of any size can use. The intended and unintended outcomes of those uses are in flux as the technologies powering development of algorithms continues at a rapid pace. Congress may want to consider looking to global policymaking processes that have prioritized education before legislation. Regulatory efforts that focus on mitigating material harms in high-risk use cases, some of which are already covered by existing regulations, appear to be resulting in an appropriate balance of responsible innovation by technology creators along with safety, security, and privacy for consumers.

19. Are you aware of your platform using surveillance advertisements to target children (anyone under the age of 18) with specific ads? If so, in your opinion, how can this be mitigated?

We are unsure what you mean by "surveillance advertisements," but TikTok does not currently allow advertisements in the TikTok under 13 experience. In the U.S. 13+ experience, TikTok places a number of targeting and content restrictions on ads that can be served to minors. For example, teens aged 13-15 cannot see personalized ads on TikTok based on their activities off TikTok. TikTok does not allow ad content and landing pages that are aimed at minors or feature

[18] https://www.tiktok.com/transparency/en-us/recommendation-system/
[19] https://newsroom.tiktok.com/en-us/learn-why-a-video-is-recommended-for-you
[20] https://www.tiktok.com/transparency/en-us/security-privacy/
[21] https://www.tiktok.com/transparency/en-us/research-api/

content which is likely to appeal to minors or to encourage minors to buy goods or services. TikTok also does not allow ad content to promote inappropriate behaviors involving minors, such as underage drinking, underage gambling, or underage smoking.

In addition, TikTok's policies require advertisers to exclude users under the age of 18 from certain types of advertising content. For example, financial services-related ads, such as those related to loans or to the exchange, management, or investment of funds, may not be targeted to users under 18 years of age. Likewise, ads for pharmaceuticals, healthcare, and medicines may not be targeted to users under 18 years of age. As another example, ads that promote dating apps may not target users under 18. The full list of content prohibited to users under 18 years of age is detailed in TikTok's Advertising Policies (https://ads.tiktok.com/help/article/tiktok-advertising-policies-ad-creatives-landing-page?).

20. Beyond surveillance advertisements, are there any other algorithmic-based practices being implemented that are particularly detrimental to children? In your opinion, how can this be mitigated?

We are unsure what you mean by "surveillance advertisements", but user safety, and especially the safety of younger users, is a primary consideration in regulating content that is shown in the For You Feed. TikTok's Community Guidelines prohibit a range of content, including nudity, violent extremism, drugs, and depiction of suicide or self-harm. We proactively remove content that violates our Community Guidelines. For example, in 2023 Q3, we removed more than 136 million violative videos globally, which accounts for just under 1% of total published videos over that period. The vast majority (96%) were removed proactively, before they were reported to us.

Some content categories that do not violate TikTok's Community Guidelines are nevertheless rendered ineligible for recommendation to the For Your Feed because they may not be suitable for all users, such as content that is determined to be disturbing and graphic. To further enhance minor safety, we age restrict certain content such as disordered eating and body image and make it unavailable to minors under the age of 18.

TikTok also has implemented measures to disperse certain content - such as content somewhat suggesting extreme dieting or fitness - that may be fine if seen occasionally but may be problematic if viewed in close or continuous succession.

21. Are you aware of any surveillance advertisements or algorithms that are used to target children, specifically to promote drugs and the sale of narcotics?

We are unsure what you mean by "surveillance advertisements", but TikTok does not allow ad content and landing pages to display or promote illegal drugs, controlled drugs, prescription drug abuse, recreational drugs, drug paraphernalia, or accessories or supplies any of such, including the use of them. Ads for pharmaceuticals, healthcare, and medicines must target 18+ users and

comply with all applicable laws and regulations of the target country. You can learn more on our website[22].

22. A Chinese government-affiliated entity owns 1% of ByteDance. On your website, you explain that this is a "common arrangement for companies operating news and information platforms in China."

It seems that the question is not whether this is a common or legal practice, but whether this is an appropriate practice. Do you consider the Chinese Government's ownership in your controlling parent company appropriate?

TikTok's parent company, ByteDance Ltd., has China-based subsidiaries including Beijing Douyin Information Services Co., Ltd. The 1% stake in that entity does not give the Chinese government any right to influence the operations of TikTok. Please see the corporate structure diagram at ByteDance.com, which illustrates the structural separation between Beijing Douyin Information Services, Co. Ltd. and all TikTok entities.

TikTok has imposed data access policies to help ensure that adequate safeguards are in place to protect personal information. If a person has no business need to access protected U.S. user data, they are not afforded such access under the terms of those policies. In addition to these safeguards, TikTok Inc.'s U.S. subsidiary, USDS, is undertaking efforts that are unprecedented among our peer group to build a secure environment for protected U.S. user data.

23. Forbes recently reported that ByteDance has started hiring healthcare and science professionals for its new AI for Drug Design and Science teams. Forbes suggests that U.S. data could be used for product feedback, marketing development, and spending patterns.

All of this data could then be fed to Chinese pharmaceutical companies, weapons manufacturers, etc. What U.S. user data from TikTok is available to ByteDance for its medical and scientific research?

This team does not use TikTok user data for its research.

24. Last March, TikTok announced that Project Texas was underway. This project is an effort to store U.S. user data in the United States through Texas based company, Oracle. What is the status of Project Texas and has TikTok finished deleting all old U.S. data from non-oracle servers?

TikTok Inc.'s U.S. subsidiary, USDS, is undertaking efforts that are unprecedented among our peer group to build a secure environment for protected U.S. user data, protect the platform from

[22] https://ads.tiktok.com/help/article/tiktok-advertising-policies-ad-creatives-landing-page-prohibited-content?lang=en

outside influence, and implement safeguards on our content recommendation and moderation tools.

Since January of 2023, new protected U.S. user data has been stored in the Oracle Cloud in an environment controlled by USDS. We've taken steps to limit to only USDS personnel access to protected U.S. user data in the Oracle environment, unless authorization is given by USDS pursuant to limited exceptions, such as for legal and compliance purposes. Traffic from the Oracle Cloud now goes through Oracle controlled gateways to prevent protected U.S. user data from being transferred to or accessed by employees of TikTok or ByteDance. TikTok continues to delete historical protected U.S. user data from our global data centers, helping to ensure that protected U.S. user data is stored in Oracle's Cloud and controlled by USDS. We have completed the first round of deletion from TikTok servers in our global data centers. In order to provide additional assurance and validation of our team's work, we have also begun independent validation efforts to confirm the comprehensiveness of the data deletions.

25. What influence does ByteDance or its employees have on recommendation algorithms? Do Chinese Communist Party Officials provide guidance on what algorithms should be in place?

Chinese Communist Party officials do not provide guidance on what algorithms TikTok should use to operate the TikTok platform. Moreover, TikTok has policies that prohibit personnel from taking any action in the course of their work for or with TikTok to advance the political agenda of a third party through the promotion, recommendation, moderation, or filtering of content.

26. Can the Chinese Communist Party lawfully require TikTok to hand over U.S. private user data through ByteDance?

TikTok does not offer the TikTok app for download in mainland China. TikTok Inc. has not been asked by the Chinese government for U.S. user data. TikTok discloses on a regular basis in its Information Requests Reports the volume and type of requests for user information received from governments and law enforcement agencies, and whether the data was disclosed or presented. See TikTok's Information Requests Report[23].

TikTok Inc.'s ultimate parent company is ByteDance Ltd., a privately-held, global holding company incorporated in the Cayman Islands and subject to the law of the Cayman Islands. ByteDance Ltd. owns many businesses, some of which operate in China. ByteDance is obligated to comply with China's laws insofar as they relate to the ByteDance business just as it is obligated to comply with the laws of the United States. Other U.S. companies doing business in China would likewise be required to comply with Chinese laws insofar as they are doing business in China.

TikTok does not make the app available for download in mainland China. Additionally, we've taken steps to limit to only USDS personnel access to protected U.S. TikTok user data in the

[23] https://www.tiktok.com/transparency/en-us/information-requests-2023-1/

Oracle environment, unless authorization is given by USDS pursuant to limited exceptions. Traffic from the Oracle Cloud now goes through Oracle controlled gateways to prevent protected U.S. TikTok user data from being transferred to or accessed by employees of TikTok or ByteDance. TikTok believes that these cutting-edge measures are unprecedented among our peer group and will provide many layers of protection for U.S. user data.

Project Texas, as we have described elsewhere, safeguards protected U.S. user data by taking steps to limit access only to USDS personnel, subject to limited exceptions.

27. At the hearing, you said that TikTok is investing $2 billion in online safety measures across your platform. How much of that funding is going to operations in the United States? What are the percentage amounts for other countries?

As a privately held company, TikTok does not disclose revenue or detailed budget breakdowns.

Senator Dick Durbin
Chair, Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **For each year from 2019 to 2023, please provide the following:**

   a. **the total number of users on your platform;**

      Discord has in excess of 150 million users globally. However we are a private company and do not publish or share monthly active user (MAU) numbers publicly. We would be happy to provide more specific U.S. figures to your office confidentially.

   b. **the total number of users under the age of 18 on your platform;**

      As stated above, Discord is a private company and does not publish or share MAU numbers publicly, no matter the demographic subset of users at issue. We would be happy to provide more specific U.S. figures to your office confidentially.

   c. **the estimated number of users under the age of 13 on your platform;**

      As Discord prohibits the usage of its platform by individuals under the age of 13, Discord does not have data responsive to this question.

   d. **the number of users of your platform under the age of 18 who were paired with, or supervised by, a parent or guardian using your Family Center tool?**

      Family Center launched in summer of 2023 so there is no data prior to that date. As of February 2, 2024, there are 15,000 parent users currently connected to 15,500 under-18 users on Discord Family Center.

   e. **your company's annual revenue;**

      Discord is a private company and does not publish or share revenue figures publicly. We'd be happy to provide these figures to your office confidentially.

   f. **your company's annual budget for trust and safety;**

      2019

      - Trust & Safety: **$2.4M** in full-time employees (FTEs) and related expenses;
        Safety Tooling: **$1.2M** (estimated)

2020

- Trust & Safety: **$4.9M** in FTEs and related expenses;
- Safety Tooling: **$3.4M**

2021

- Trust & Safety: **$9.4M** in FTEs and related expenses;
- Safety Tooling: **$24.3M**

2022

- Trust & Safety: **$18.4M** in FTEs and related expenses;
- Safety Tooling: **$36.3M**

2023

- Trust & Safety: **$29.6M** in FTEs and related expenses;
- Safety Tooling: **$60M**

The financial information provided above does not include cross-functional efforts that contribute to trust and safety, including things like engineering efforts, marketing, policy, legal support, and more. Discord is unable to estimate those numbers prior to 2023 but in that year those efforts accounted for ***$30.8 million dollars*** *in direct expenses.*

g. your company's annual budget to address online child sexual exploitation;

Discord's budget is not organized in a manner which allows it to determine funding dedicated to solely addressing online child sexual exploitation, but substantial resources from Trust & Safety, Engineer, Legal, Policy, and other teams work towards addressing this issue.

h. the total number of employees working to address trust and safety;

- **2019:** 22 FTEs within Trust & Safety
- **2020:** 37 FTEs within Trust & Safety
- **2021:** 81 FTEs within Trust & Safety
- **2022:** 79 FTEs within Trust & Safety
- **2023:** 90 FTEs within Trust & Safety

The information provided above does not include employees on cross-functional teams that contribute to trust and safety, including engineering, marketing, policy, legal support, and more. Discord is unable to estimate those numbers prior to 2023 but in that year approximately ***75 additional employees*** *contributed to efforts to address trust and safety. It also does not include external partner agents. That number grew from 10 additional external partner agents in 2019 to* ***approximately 250 in 2023****.*

i. **the total number of employees working to address online child sexual exploitation.**

Discord has more than 20 percent of its total Trust and Safety employees dedicated to Minor Safety efforts, and nearly 40 percent of its external partners involved with child safety reports. Discord has 18 full-time employees and 158 external partners directly involved in these issues.

2. **How did your company determine that 13 was the appropriate age for a child to begin using your platform?**

We allow teens to access our services in accordance with local requirements and federal law.

3. **What legal obligation does your company have in the United States to ensure that your platforms are safe for children before they are launched?**

In the United States, Discord is subject to a variety of federal and state laws that address children's safety. On a federal level, Discord must comply with the Children's Online Privacy Protection Act (COPPA). States have and continue to create legal and regulatory frameworks to bolster children's privacy and online safety.

4. **For users under the age of 18,**

a. **what are the default privacy settings for their accounts?**

User profiles are not public in the ways one might experience on other platforms. To start, one cannot begin searching for users and see a list of possible matches to add as a friend. Instead, one must know and then specifically search for the username associated with the account they would like to add as a friend. This layer of searchability protection applies to all account types.

The full range of user settings are available and can be customized via the user settings menu. For example, while users can see one another's usernames when in the same server, Discord provides each user control over who can try to add them as friends. A user can apply these controls by toggling the options under the section titled "Who can send you a friend request" of the "Friend Request" option in the user settings. There, a user has three categories by which they can limit friend requests: "Everyone;" "Friends of Friends;" and, "Server Members." Users can also apply controls that restrict who can send them DMs, and by default teens can only be DM'd by their friends and members of the communities they have chosen to join.

In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. One of the tools announced as a part of this new initiative is Sensitive Content Filters, which automatically blur

potentially sensitive media sent to teens in direct messages (DMs), group direct messages (GDMs), and in servers. Blurring is enabled by default for teen users both in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users. Adult users can also choose to opt-in to this feature by changing their Privacy & Safety Settings.

Last fall, Discord also launched Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn that leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify a teen user about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

b. what limitations are placed by default on content these users can access, content that will be directed toward them, and individuals they can communicate with?

Discord's Community Guidelines require that all adult content posted to Discord be kept behind an age-restricted gate that prevents access by users under 18 years of age. If a server is organized around age-restricted themes or if the majority of the server's content is focused on adult content, the server must be classified as age-restricted. The Guidelines' restriction on adult content to age-restricted spaces applies to both servers and channels. Users under 18 years of age are not able to join or view the content in age-restricted servers and channels.

As explained above, Sensitive Content Filters and Safety Alerts for Senders are on by default for teen users. Safety Alerts in Chat, our most recent safety feature, will be rolled out in the coming weeks and will be enabled by default for teens.

c. can they change their default settings without the awareness of their parent or guardian, or without the consent of their parent or guardian?

Users under 18 years of age cannot alter Discord's settings in order to access content in age-restricted servers or channels. Users under 18 years of age also cannot disable Sensitive Content Filters or the Safety Alerts for Senders features described above.

d. in 2023, how many changed their default settings?

In 2023, approximately 278,000 U.S. monthly active users under the age of 18 changed any default privacy setting at some point during the year; as stated above, however, users under 18 years of age cannot alter Discord's settings in order to access content in age-restricted servers or channels nor can users under 18 years of age disable Sensitive Content Filters or the Safety Alerts for Senders features.

5. If the default settings are different for users aged 16 and 17 than they are for users under the age of 16, please explain why your company takes this position, how this position was developed, and whether any company personnel voiced objections to or raised concerns about this position.

Default settings for users aged 16 and 17 are no different for users under the age of 16.

6. What studies, research, summaries, or data does your company have reflecting the efficacy of its parental controls and child safety measures? Please provide these studies, research, summaries, or data.

Discord tracks the efficacy of its automated child sexual abuse material (CSAM) detection systems at a daily cadence. This includes examining the detection and precision of models, the outcomes, the number of false positives, the overall number of images detected by the model, and the proactivity rate. Discord uses this information to inform future improvements to its detection models and safety tools, review the efficacy of its interventions and existing tools, and gain insight into their current state of operation.

For example, in early Q1 2023, Discord's PhotoDNA false positive rate was approximately 45 percent. Through analysis enabled by these tools, that rate has fallen to approximately 20 percent. This improvement was driven both by examination of data related to the efficacy of Discord's CSAM detection and intervention tools, and investment in and implementation of new tools. For example, Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. As a result, our internal hash database is now able to identify more CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online. By leveraging these tools, Discord's model had an 87 percent increase in recall, i.e. proactively-detected confirmed positive CSAM.

Further improvements are coming shortly, such as the ability to ignore repeated false positives to prevent a Discord agent from needing to examine a known-false positive image more than once, which will improve precision even further.

7. **Concerning international law,**

   a. **what steps have your company and its subsidiaries taken to comply with the European Union's Digital Services Act?**

      Discord has undertaken work in order to meet our obligations under the DSA, including:

      - updating our new Warning System hub—a feature available to users worldwide—to allow EU users to login and view their account standing, even when the account is suspended, and adding more context to our action notices, including the content that violated our policies (as long as providing that additional context does not harm Discord or others)t;
      - updating our Safety Center with information about how to submit DSA reports, appeal enforcement decisions, and more;
      - created a Reports & Appeal process for DSA-eligible reports; and
      - added dedicated pathways for DSA-required government requests and trusted flagger reports to our existing law enforcement portal, among other steps.

   b. **what steps has your company and its subsidiaries taken to comply with the United Kingdom's Online Safety Act?**

      Discord has followed closely the drafting and passage of the United Kingdom's Online Safety Act (UK OSA). The UK OSA's requirements are not yet fully defined. Discord is currently reviewing the draft codes as they become available and is engaging with the relevant UK regulator.

   c. **what steps has your company and its subsidiaries taken to comply with Australia's Online Safety Act?**

      Discord has evaluated its existing practices against the obligations under Australia's OSA and the existing codes and standards, and we have published information for Australian users about our obligations under the law.

   d. **if those laws create a safer, healthier online experience for kids on your platforms, do you commit to implement these changes in the United States? If not, why not?**

      We view user safety not only as our responsibility but absolutely essential to fulfilling our mission and making sure Discord continues to be the best place to hang out and have fun with friends. We will evaluate and implement practices

globally that further this work. For example, we plan to evaluate the features we built to meet our obligations under the DSA to identify those that improve safety for all of our users.

8. **This past summer, NBC News identified 242 publicly listed Discord servers created in the previous month that appeared to openly market sexually explicit content of minors, including at least 15 that directly appealed to teens by claiming to be sexual communities for minors. Some of these communities had over 1,500 members.**

   Additionally, your platform allows for text, audio, and video chat, facilitating the creation and sharing of CSAM.

   What steps is your company taking to identify and remove these servers, users, and CSAM content? How does your company measure the efficacy of those steps?

   In addition to the hash matching described in our other answers, as well as our use of CLIP to detect unknown and AI-generated CSAM, Discord uses internal tools (including machine learning techniques) and works with industry partners (including peer companies, non-profits, and researchers) to detect networks of actors engaging in CSAM distribution. This approach, leveraging technology and human expertise, allows us to better identify and remove spaces and users on Discord related to CSAM content.

   Discord also relies on users reporting violative material as part of our response to CSAM across the platform.

   Users who upload abuse material of children onto Discord are reported to the National Center for Missing and Exploited Children (NCMEC) and removed from the service. We deeply value our partnership with NCMEC and their efforts to ensure that grooming and endangerment cases are quickly escalated to law enforcement.

   In addition, Discord works with leading groups and partner organizations to assist in the detection of CSAM and the improvement of our internal policies and processes, including NCMEC, the Family Online Safety Institute and the Technology Coalition (a group of companies working together to end online child sexual exploitation and abuse), among others. We are also a frequent sponsor of events dedicated to increasing awareness of, and action on, child safety issues such as the annual Dallas Crimes Against Children Conference.

   As part of our Safer Internet Day efforts in 2023 we partnered with NoFiltr, the youth-targeted online safety program by Thorn, to develop online safety resources for young people. These resources included an interactive quiz, online fortune teller, and a social media campaign aimed at supporting healthy digital habits and online safety.

   Finally, we invest heavily in advanced tooling and education so parents know how our service works and understand the controls that can contribute to creating a positive and safe experience on Discord for their children. Discord also continues to innovate our

approach to child safety. As described above, our recently announced Teen Safety Assist initiative contains new tools to keep teens safe on Discord: Sensitive Content Filters and Safety Alerts for Senders are already in place and working to make Discord a safer place for teens; Safety Alerts in Chat will be rolled out in the coming weeks.

9. **YouTube, TikTok, and Twitch all use technology to monitor livestreaming for sexual abuse. Discord does not. In a statement, Discord said that, "Discord does not monitor livestream content," because spending money to monitor that content would be "at the detriment of other Discord safety programs."**

    a. **How would monitoring livestreams on Discord be detrimental to other Discord safety programs?**

    Discord does not have broadcast livestream capabilities; Discord users are only able to livestream within the communities of which they are a member. This inherently limits the distribution of the content streamed, and Discord therefore does not monitor or record livestream content or voice chats. Discord has instead prioritized resources into other forms of CSAM detection as described in response to other questions. Monitoring livestreaming at the scale Discord operates would require that we redirect resources away from these other critical child safety functionalities.

    b. **How does Discord choose which aspects of its platform to monitor for child abuse and which to ignore?**

    Discord takes the detection, removal, and reporting to NCMEC of CSAM seriously, and as a mid-sized company that has grown exponentially over the last several years, we have had to rigorously prioritize our resourcing related to safety. Our primary goal has been to build a strong, research-driven and values-based foundation on which we can continue to improve with additional tooling and technology.

    Discord also uses Machine Learning (ML) in targeted ways to help identify potential malicious use at the server, channel, and user registration level to identify potential bad actors or groups for further human review. As part of Discord's efforts to improve ML-based moderation, in 2021, Discord purchased Sentropy, a company that builds ML and Artificial Intelligence tools to help companies moderate disruptive behavior on their platforms. Discord also uses industry-standard hashing and matching technology, including PhotoDNA, for other types of illegal conduct, such as CSAM.

10. **A few weeks ago, the Mayor of Highland Park, Illinois, shared that Discord has been linked to several perpetrators of mass shootings across the county. The individual who shot and killed seven people and wounded 47 others at a Fourth of July parade in Highland Park had his own Discord server that promoted violent imagery. The perpetrator of the Buffalo, New York mass shooting used Discord to**

make a to-do list for his attack. Last month, a 17-year-old who fatally shot two and injured six others at his school in Iowa used Discord to chat about school shootings and wrote about "gearing up" and assembling his guns just moments before he opened fire in his school's cafeteria.

What steps has your company taken to ensure Discord identifies potentially violent individuals and reports them to law enforcement?

Discord has specialist teams trained in evaluating and acting upon certain types of high-harm conduct, including Violent Extremism, Exploitative and Child Sexual Abuse Materials, and Cybercrime. These teams engage in proactive moderation using a combination of human interventions and machine learning. These proactive efforts are detailed in Discord's Transparency Reports and in additional materials providing an overview of the Trust & Safety team and team-by-team breakdowns. They also receive regular updates from law enforcement, the Global Internet Forum to Counter Terrorism (GIFCT) and other organizations, participate in events and conferences run by academic experts, review resource articles and books, and study language and phrases used by extremist groups. All of these efforts enable specialist moderation teams to better interpret the material they are reviewing and make informed choices.

Discord also uses Machine Learning (ML) in targeted ways to help identify potential malicious use at the server, channel, and user registration level to identify potential bad actors or groups for further human review. As part of Discord's efforts to improve ML-based moderation, in 2021, Discord purchased Sentropy, a company that builds ML and Artificial Intelligence tools to help companies moderate disruptive behavior on their platforms. Discord also uses industry-standard hashing and matching technology, including PhotoDNA, for other types of illegal conduct, such as CSAM.

11. **You testified that Discord does not encrypt messages on its platform. You stated that, "This is a choice we made." You further explained that, "We don't believe we can fulfill our safety obligations if the text messages of teens are fully encrypted, because encryption would block our ability to investigate a serious situation when appropriate report to law enforcement."**

Yet, Discord published a blog post last summer in which it announced that it was "beginning to experiment with new encryption protocols and technologies for voice and video calls on Discord."

Do you commit to keeping communications involving users under the age of 18 unencrypted?

Text messages on Discord are not end-to-end encrypted and we commit to keeping teen users' text messages unencrypted. The features Discord deploys to help keep teens safe on our service, such as our Teen Safety Assist feature described in detail above, are designed to identify conversations with teen users that could present a risk to their safety,

and the effectiveness of our teen safety features would be undermined by encrypting the messages of teen users.

We believe user expectations for the privacy of voice and video calls are higher than for text messages. Accordingly, to meet user expectations, we are evaluating implementation of end-to-end encryption for voice and video traffic on our platform.

12. **According to Discord's website, it takes a "proactive and automated approach to safety" only on servers with more than 200 members. Smaller servers rely on server owners and community moderators to define and enforce norms of behavior.**

Indeed, in a September 2021 interview, Mr. Citron described how most people engage with Discord in smaller groups of 10 or 20 people and that "in those kind of small spaces…we leave it to…people to decide…what…is acceptable…in their space."

Real world experience shows the risk of this approach. Last summer, for example, NBC News reported that adults had used hidden communities and chat rooms on Discord to "groom children before abducting them, trade child sexual exploitation material…and extort minors whom they trick into sending nude images."

How do you defend an approach to safety that relies on groups of fewer than 200 sexual predators to report themselves for things like grooming, the trading of CSAM, and sextortion?

The above quote from my September 2021 CNBC interview omits relevant information that I shared about Discord's approach to safety, which does not rely solely upon user-generated reports. Discord leverages a mix of both proactive detection and reactive measures, supported by a range of human-powered and technical solutions.

During the interview, I explained that "in those kind of small spaces … we leave it to … people to decide … what … is acceptable … in their space *in the context of our Community Guidelines*." I further explained that "if people encounter content that … let's say that the moderators in a space aren't deleting or is violating our Guidelines, we do have a Trust & Safety team, full time employees that people can escalate issues to, and because Discord is not end-to-end encrypted, although we don't proactively read people's messages, if people forward messages to or report content to our Trust & Safety team, we will go investigate and we will action communities that are violating our Guidelines."

It bears repeating that Discord has a zero-tolerance policy for content or conduct that endangers or sexualizes children. Our Community Guidelines prohibit CSAM and child sexualization. We do not allow CSAM on Discord, including AI-generated, photorealistic CSAM. We proactively scan images uploaded to our service—no matter the size of the server in which a user seeks to post an image or the number of users participating in the conversation—in order to detect, remove, and report CSAM content and perpetrators to NCMEC, which subsequently works with local law enforcement to take appropriate

action. These proactive measures are deployed in smaller servers and direct messages just as they are deployed in large group spaces on our service. In addition to our use of industry standard tools to detect and remove known CSAM, Discord invests in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. Additionally, as described in response to a previous question, Discord uses ML models to harness metadata from known communities that violate our policies to detect and remove similar communities.

Discord also has a zero-tolerance policy for inappropriate sexual conduct with children, meaning inappropriate sexual contact between adults and teens on the service, with special attention given to predatory behaviors such as online enticement and the sexual extortion of children, commonly referred to as "sextortion." We also invest in new ways to detect text-based sexual exploitation and extortion of minors, including working collaboratively with Thorn to develop a machine learning model that will detect when a minor is in communication with someone actively trying to exploit or extort them.

Moreover, in the time since I participated in that interview, Discord has created new features specifically designed to keep teens safe on the platform—features that are deployed in any space in which a teen has conversations on Discord, no matter the size of the server or how many users are participating in the conversation. Our new safety initiative, Teen Safety Assist, makes new tools available to proactively protect teens and offer tips to make their time on Discord safer. In addition to the tools described below, we are continuing to invest in safety technology and we look forward to adding even more protective messaging features for teen users in the future.

a. Safety Alerts on Senders were designed in collaboration with the child safety non-profit Thorn to alert teen users on Discord when a conversation is potentially unwanted. This new feature helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

b. Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

c. Sensitive Content Filters automatically blur potentially sensitive media sent to teens in DMs, GDMs, and in servers. Blurring is enabled by default for teen users in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users.

d. The Family Center was designed to provide parents and guardians with tools to better understand how their teens use our service. Our Family Center allows parents and guardians to get insights into the communities their teen children have joined, including which users their teens have recently added as friends, which users their teen has messaged or called in DMs or GDMs, and which servers their teen has joined or participated in. The Family Center also allows parents and guardians to get a weekly email with a high-level summary of the time their teen spends on Discord. The Family Center is designed not just to provide parents and guardians with greater insight into how their teen uses Discord, but to also help parents start conversations with their teen about online safety.

Senator Lindsey O. Graham
Ranking Member, Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **Do you support S. 1207, the bipartisan EARN IT Act? Why or why not?**

   Discord is aligned with the goals of the EARN IT Act—to curb the availability and distribution of child sexual abuse material (CSAM) online. However, Discord is concerned that the imposition of broad liability could inadvertently result in the over-moderation of content by online services, including the blocking and suppression of otherwise constitutionally protected speech. We look forward to working with the Committee to address these and other concerns.

2. **What measures are you taking to prevent and address sextortion, including financial sextortion, on your companies' platforms?**

   As with our efforts to prevent and disrupt the dissemination of CSAM, Discord's approach to child safety more broadly leverages a mix of both proactive detection and reactive measures, supported by a range of human-powered and technical solutions. Examples relevant to the issue of sextortion include Discord's Safety Alerts for teen users, which are designed to detect unwanted interactions via text message and better equip teens to safely navigate such situations.

   In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. Among the tools announced as a part of this new initiative are Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

   On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, which leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

Additionally, Discord has also invested in staffing a dedicated team housed under our Trust & Safety Minor Exploitative Content team whose sole focus is addressing high-harm sextortion activity on the platform.

a. **What methods are in place to detect and disrupt this type of abuse in real time?**

Discord's Safety Alerts on Senders and Safety Alerts in Chat operate in real time.

3. **Please provide the committee statistics on how long it takes your company to respond to various types of legal process from law enforcement?**

Discord reviews and responds to all legal requests from law enforcement in a timely manner. We comply with valid legal process before any indicated due date and communicate with law enforcement on the status of their request. For emergency disclosure requests, Discord strives to respond within 15 minutes and to have a full resolution within an hour of receipt.

4. **Do you notify your users when law enforcement serves subpoenas/summons for subscriber information and specifically requests not to notify the subscriber/user?**

As an electronic communication service provider, Discord strictly complies with the provisions of the Electronic Communications Privacy Act (ECPA), and discloses information to law enforcement in accordance with the provisions of the statute. With regards to user notice, Discord would not provide a user with notice when we receive an order prohibiting us from doing so. Moreover, if the legal process pertains to child sexual exploitation or the danger of death or serious bodily injury, Discord does not provide notice to help avoid a harmful or dangerous outcome resulting from the notice.

Discord does not notify users when it receives a 90-day preservation order under the provisions of existing law. See 18 USC 2703(f). Thus, after sending a preservation order to Discord, law enforcement always has 90 days to secure proper legal process, including a non-disclosure order, while relevant materials are preserved.

Additionally, if Discord receives legal process without a prior preservation order and without a non-disclosure order, even in cases where Discord intends to provide notice to a user, we would first alert law enforcement of this plan and allow law enforcement to withdraw the subpoena or delay its enforcement until a non-disclosure order can be obtained. This policy effectively ensures that users are not notified of legal process in circumstances where law enforcement wants to prevent such notification. More information on Discord's practices can be found in this article: How Discord Works With Law Enforcement.

a. **If you notify the subscriber, how long do you wait until notification goes out?**

Discord's policy does not mandate a waiting period prior to user notification;

however, Discord does not provide a user with notice when we receive a confirmation from law enforcement that we are prohibited from doing so.

b. Are you aware that by notifying the subscriber about a law enforcement subpoena for their subscriber information that you are jeopardizing critical evidence that could be erased before law enforcement can serve warrants?

Discord preserves the evidence requested about subscribers before possibly notifying them of any law enforcement process. Thus, we do not jeopardize critical evidence on our systems. Where law enforcement believes that notification in a particular case could cause the type of adverse effect you describe, law enforcement can, and often does, serve Discord with a non-disclosure order requiring that notification be delayed. This is the appropriate method by which non-disclosure orders should be obtained. Discord is one of many providers, not limited to technology companies, who have adopted policies to provide subscribers with notice when their material is sought to be obtained by law enforcement.

c. Would your company agree to a 90-day non-disclosure to subscribers to allow law enforcement ample time to secure proper legal process?

Discord does not notify users when it receives a 90-day preservation order under the provisions of existing law. See 18 USC 2703(f). Thus, after sending a preservation order to Discord, law enforcement always has 90 days to secure proper legal process, including a non-disclosure order, while the relevant materials are preserved. Similarly, when Discord receives legal process without a preservation order, even in cases where Discord intends to provide notice to a subscriber, we would first alert law enforcement of this plan and allow law enforcement to withdraw the subpoena or delay its enforcement until a non-disclosure order can be obtained. This policy effectively ensures that subscribers are not notified of legal process in circumstances where law enforcement wants to prevent such notification.

d. Do you actively seek out and incorporate feedback and insight from survivors of online sexual exploitation to improve your trust and safety policies and practices and to prevent and disrupt child sexual abuse material (CSAM) production and distribution on your platform? Can you provide examples?

Discord's policies and practices are driven by expert research that draws from diverse perspectives, including survivors. For example, we collaborated with the researchers from Thorn to ensure we were taking a research-backed approach to designing the Teen Safety Assist feature to ensure that it reflected the lived experience of teens.

Additionally, through our participation in Tech Coalition and other convenings led

by groups like WeProtect Global Alliance and the Internet Watch Foundation, we are committed to constantly learning and evolving our approach to better support young people and those who may be victims of online sexual abuse. For example, earlier this year three Discord employees attended the Multi-Stakeholder Forum to Combat the Financial Sextortion of Young People, where discussions about developing safety tools and approaches centered on the experiences of young people and victims. Discord also participated in the 2022 WeProtect Global Alliance convening which featured a presentation from the Brave Movement, a survivor-centered global movement fighting to end childhood sexual violence. Discord is a proud sponsor and supporter of the annual Crimes Against Children Conference, organized by the Dallas Children's Advocacy Center. This annual conference brings together key stakeholders from industry, law enforcement, and advocacy to provide specialized training and knowledge sharing in various critical fields involved in crimes against children. Discord also plans to participate in the National Center for Missing and Exploited Children's (NCMEC) upcoming CyberTipline Roundtable, where stakeholders gather for collaborative discussion of potential improvements to the reporting process.

e. If not, please explain.

Please see the above answer.

5. **During our hearing, you testified that you collaborate with parents and parent organizations to create mechanisms to keep children safe online. Please elaborate and cite examples of your company's work with non-employee parents and parent organizations.**

Discord's work with parents and parent-led organizations is vital to our child safety efforts.

Since 2022, Discord has been a sponsor of the National PTA. Discord collaborates with National PTA on digital safety resources and events designed to elevate youth voices and foster shared understanding among teens, caregivers and educators. The PTA Connected: Build Up and Belong program, sponsored by Discord, facilitates technology discussions between teens and their parents and caregivers about online safety topics. In 2023, 25 local PTAs were selected to receive $1,250 each, sponsored by Discord, to host a PTA Connected: Build Up and Belong program. Discord also sponsored grants for this program in 2022.

Discord is also a member of the Family Online Safety Institute (FOSI). FOSI develops resources to provide parents and guardians with the tools they need to confidently navigate the online world with their families.

Discord also supports ConnectSafely, a non-profit organization that educates people about online safety, privacy, security, and digital wellness. ConnectSafely has developed research-based safety tips, as well as in-depth guides and "quick-guides" for parents,

educators, youth, and policymakers. Discord worked with ConnectSafely to develop a Parent's Guide to Discord.

In recognition of the unique safety challenges that accompany teens' use of the internet, Discord has also created special features designed to give parents and teens even more control over their experience on Discord. The Family Center was designed to provide parents and guardians with tools to better understand how their teens use our service. Our Family Center allows parents to get insights into the communities their teen children have joined, including which users their teens have recently added as friends, which users their teen has messaged or called in DMs or GDMs, and which servers their teen has joined or participated in. The Family Center also allows parents to get a weekly email with a high-level summary of the time their teen spends on Discord. The Family Center is designed not just to provide parents and guardians with greater insight into how their teen uses Discord, but to also help parents start conversations with their teen about online safety.

6. **Why does your company have the age limit of 13 years old for a user to sign up for an account?**

   We allow teens to access our services in accordance with local requirements and federal law.

   a. **Why not younger or older?**

      Please see the answer to the preceding question.

7. **How many minors use your platform? How much money does your company make annually from these minors?**

   There are approximately 2.7 million monthly active users under the age of 18 in the U.S. Discord is unable to answer the second part of the question as age data is segregated from revenue data and cannot be aggregated together.

8. **What percentage of your employees work on trust and safety and how much money does your company invest annually in trust and safety?**

   More than 15 percent of our employees work on safety at Discord. By way of context, that means our Safety team is larger than our Marketing team. In addition to the significant investment we have made in our full-time employees working on these issues, we also work closely with external partners to supplement our own work on critical safety issues. In 2023, Discord invested **$29.6M** in full-time employees and related expenses, **$60M** in safety tooling, and **$30.8M** in cross-functional efforts that contributed to trust and safety, such as related engineering, marketing, policy, and legal support.

9. **It is sometimes challenging for law enforcement conducting criminal investigations to determine the true identity of a person behind a name on social media or other**

online platforms, and whether an online identity is an actual person. What are you doing to validate the true identity of users – or the fact that a user is a human – when they create an account on your platforms?

To access the services on an ongoing basis, Discord requires users to submit an email address and/or a phone number upon account registration.

Discord also imposes a CAPTCHA when a login comes from a new IP address range—reducing the efficiency of automated systems attempting to access and use the platform.

10. **Is your company using safety technology to detect and prevent live video child sexual abuse on your platforms and apps that allow users to stream or share live video? If not, please explain.**

    Discord does not have broadcast livestream capabilities; Discord users are only able to livestream within the communities of which they are a member. This inherently limits the distribution of the content streamed, and Discord therefore does not monitor or record livestream content or voice chats. Discord has instead prioritized resources into other forms of CSAM detection as described in response to other questions. Monitoring livestreaming at the scale Discord operates would require that we redirect resources away from these other critical child safety functionalities.

    a. **Has your company tested that or similar technology? If not, are you developing similar technology to address child sexual abuse in live video?**

       We have not tested or developed technology with regard to live streaming or live videos, however Discord regularly evaluates new technologies and best practices relating to audio and video content moderation.

11. **How are you measuring if your trust and safety policies, practices, and tools are effective in protecting children from sexual abuse and exploitation on your platform?**

    Discord tracks the efficacy of its automated CSAM detection systems at a daily cadence. This includes examining the detection and precision of models, the outcomes, the number of false positives, the overall number of images detected by the model, outcomes of, and the proactivity rate. Discord uses this information to inform future improvements to its detection models and safety tools, review the efficacy of its interventions and existing tools, and gain insight into their current state of operation.

    a. **What specific metrics or key performance indicators do you use?**

       In measuring the effectiveness of our trust and safety policies, practices, and tools, Discord evaluates the detection and precision of models, outcomes, the number of false positives, the overall number of images detected by the model, and the

proactivity rate, among other factors.

12. Is your company using language analysis tools to detect grooming activities? If not, please explain.

On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature developed in collaboration with Thorn, a leading non-profit in the minor safety space, which leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

a. What investments will your company make to develop new or improve existing tools?

Discord invests in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and thus do not appear in available hash databases. We also invest in new ways to detect the exploitation of minors (commonly referred to as "grooming").

- **Novel CSAM:** Recently, Discord collaborated with industry peers to build and implement a visual safety technology to detect unknown CSAM and AI-generated CSAM with positive results. Effective tools that keep children safe online should be industry standards, not a means for companies to gain a competitive advantage. In recognition of that, as of October we have made this technology open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online.

- **Exploitation of Minors:** We worked collaboratively with Thorn, a leading non-profit in the minor safety space, to develop Safety Alerts in Chat, a new feature that leverages machine learning algorithms to analyze messages between adults and teenagers in order to identify messages that may be indicative of grooming behaviors. This technology continues to be developed for broader industry use and further demonstrates our ongoing commitment not just to make Discord—but also the broader communications ecosystem—safer for kids.

13. What resources have you developed for victims and survivors of abuse on your platforms?

To support our users, Discord partners with Crisis Textline, a non-profit that provides

24/7 text-based mental health support and crisis intervention via trained volunteer Crisis Counselors.

We also expanded our partnership with INHOPE, a global network combatting CSAM online. Through this partnership, we are better able to collaborate with the hotlines globally working to address this important issue.

14. What is your response to requests for content removal from CSAM survivors and other members of the public?

Reports concerning the highest harm material, including reports of CSAM or inappropriate contact with a minor, are prioritized for review and enforcement. For individuals working with law enforcement agencies, Discord processes removal and takedown requests through an escalated pathway to ensure a rapid response.

15. In discussions with your leadership team, there was mention of particular safety measures on large servers. Isn't it more likely for nefarious activity to go on in smaller servers? How do you monitor and scrutinize the smaller servers?

Discord's core safety measures apply across the entire surface of our platform, in large servers and smaller spaces alike. We proactively scan images uploaded to our service—no matter the size of the server in which a user seeks to post an image or the number of users participating in the conversation—in order to detect and report CSAM content and perpetrators to NCMEC, which subsequently works with local law enforcement to take appropriate action. These proactive measures are deployed in smaller servers and direct messages just as they are deployed in large group spaces on our service. In addition to our use of industry standard tools to detect and remove known CSAM, Discord invests in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. Additionally, Discord uses ML models to harness metadata from known communities that violate our policies to detect and remove similar communities.

Discord has also created new features specifically designed to keep teens safe on the platform—features that are deployed in any space in which a teen has conversations on Discord, no matter the size of the server or how many users are participating in the conversation. Our new safety initiative, Teen Safety Assist, makes new tools available to proactively protect teens and offer tips to make their time on Discord safer. In addition to the tools described below, we are continuing to invest in safety technology and we look forward to adding even more protective messaging features for teen users in the future.

- Discord Safety Alerts were designed in collaboration with the child safety non-profit Thorn to alert teen users on Discord when a conversation is potentially unwanted. This new feature helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from

strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

- Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

- Sensitive Content Filters automatically blur potentially sensitive media sent to teens in DMs, GDMs, and in servers. Blurring is enabled by default for teen users in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users.

16. What voluntary hash-sharing or other information sharing initiatives does Discord participate in to help combat child sexual exploitation?

Discord uses PhotoDNA to proactively scan images uploaded to our service in order to detect and report CSAM content and perpetrators to NCMEC, which subsequently works with local law enforcement to take appropriate action. In addition to hash matching, Discord uses internally developed tools (including machine learning techniques) and works with industry partners (including peer companies, non-profits, and researchers) to detect CSAM distribution.

In addition to our use of industry standard tools to detect and remove known CSAM, Discord invests in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online. By leveraging these tools, Discord's model had an 87 percent increase in recall, i.e., proactively-detected confirmed positive CSAM.

We are also investing in new ways to detect the exploitation of minors (commonly referred to as grooming), such as the Teen Safety Assist features described above.

As part of our extensive fight against CSAM, we continued to deepen our partnership with the Tech Coalition. One of the most exciting parts of our partnership with the Tech

Coalition this year was the launch of Lantern, a first-of-its-kind signal sharing program for companies to enhance and strengthen how we detect attempts to sexually exploit and abuse children and teens online. Discord joined other industry partners like Google, Meta,, Roblox, Snap, and others to share signals about activity we detect that violates our policies and utilize signals shared by other companies to surface violating activity on their platforms.

Discord also works closely with NCMEC, utilizing their hash database in our internal models as well as developing technology to detect novel CSAM and opening that technology up to other companies through Project Lantern.

17. Why does Discord not participate in NCMEC's "Take It Down" program to help stop the sharing of and remove nude and sexually explicit photos of minors?

Discord does not yet participate in "Take It Down" because our work to integrate the program's hashing technology is not yet complete. Discord intends to complete that work, enabling Discord's participation in the program.

Senator Sheldon Whitehouse
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **What exemptions from the protections of Section 230 would your company be willing to accept?**

   Section 230's protections from liability are not absolute. For example, all federal criminal liability is already exempted from Section 230, as are intellectual property violations. To the extent that Congress contemplates further limiting Section 230's liability protections, it should take care not to stifle innovation or to inadvertently disincentivize platforms from moderating a broad range of potentially harmful content online.

2. **Is it your belief that your company should enjoy absolute immunity under Section 230 from suits like *Doe v. Twitter*, No. 21-CV-00485-JCS, 2023 WL 8568911 (N.D. Cal. Dec. 11, 2023), no matter the extent of your company's failure to remove reported child sexual abuse material from the platform or to stop its distribution?**

   We do not believe that companies have "absolute immunity" under Section 230. There are already existing limits in the statute, including for violations of federal criminal law or in cases where interactive computer services create the content in question. The case law also is continuing to evolve.

   Section 230 supports our ability to provide a service that enables community, expression, and connection. It also empowers us to take responsible action to remove harmful content, to enforce our policies, and to develop innovative approaches to online safety. Section 230 is also key to economic growth—allowing innovative services to flourish in the United States and in turn creating new jobs and opportunities for Americans.

Senator Chris Coons
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **During the hearing, I asked the five witnesses whether the platform they represented publicly discloses "an estimate of the total amount of content—not a percentage of the overall…but the total amount of content on your platform—that violates" the platform's "policies prohibiting content about suicide or self-harm." I also asked if each platform "report[s] the total number of views that self-harm or suicide-promoting content that violates that policy gets on [each] platform." In response to these questions, you testified while under oath "Yes, we do."**

   After reviewing Discord Inc.'s ("Discord") most recent transparency report from December 2023, it appears that your testimony was misleading. First, Discord's transparency report does not disclose an estimate of the total amount of content on Discord's platform that violates the company's suicide and self-harm policy. Second, the transparency report does not disclose the estimated number of views that this violating content on the platform receives. These exclusions stand in direct contrast to your sworn testimony on January 31.

   a. **Please provide the specific citation to where Discord publicly discloses an estimate of the total amount of content on the platform that violates Discord's suicide and self-harm policy.**

      Thank you for this opportunity to clarify my testimony. I misunderstood the question posed during the hearing and misspoke. My answer referenced the section in Discord's quarterly transparency reports titled "Self Harm Concerns" that describes our work to prevent the normalization, promotion, and encouragement of self-harm and also discloses the enforcement actions we take on accounts and servers found to violate our policy prohibiting the glorification or promotion of suicide or self-harm.

      The enforcement data Discord reports on a quarterly basis represent actions our platform takes on content found to have violated our platform policies. The data disclosed is not an estimate of the total amount of violative content, but the actual amount discovered and actioned appropriately. In the first quarter of 2023, for example, 1,294 accounts and 540 servers were removed for self-harm concerns.

   b. **Please provide the specific citation to where Discord publicly discloses an estimate of the total number of views of content that violates Discord's suicide and self-harm policy.**

      Again, thank you for the opportunity to clarify my testimony. The enforcement data represent actions the platform takes on content found to have violated our platform policies, but Discord does not track the number of times a user views a

piece of content on the platform or impressions on content, regardless of whether the content was deemed violative of our Community Guidelines.

Tracking impressions on individual pieces of content is a functionality typically built to sustain business models that incorporate algorithms that promote content to users and/or rely on the sale of advertising, both of which are common to traditional social media platforms. As a chat and messaging app, Discord has not sought to track impressions or build other common social media features like news feeds.

c. If Discord does not disclose these metrics, why not?

As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, regardless of whether the content was deemed violative of our Community Guidelines.

d. Does Discord measure these metrics? If not, why not?

No. As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, regardless of whether the content was deemed violative of our Community Guidelines.

2. Discord has previously reported how much content it removes under the platform's suicide and self-harm policy.

a. For content that has been removed, does Discord measure how many views that content received prior to being removed? If not, why not?

No. As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, regardless of whether the content was deemed violative of our Community Guidelines.

b. For content that has been removed, does Discord disclose how many views that content received prior to being removed? If so, please provide a specific citation to where Discord discloses that information. If not, why not?

No. As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, regardless of whether the content was deemed violative of our Community Guidelines.

c. Please provide an estimate of the number of views content that was removed under this policy received in January 2024.

We are unable to estimate the number of views an individual piece of content receives on Discord, regardless of whether the content was deemed violative of our Community Guidelines.

d. For content that has been removed, does Discord measure demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If not, why not?

No. As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, nor does Discord track the demographic data of users who may view an individual piece of content, regardless of whether the content was deemed violative of our Community Guidelines or the demographics of users who may have viewed the content.

e. For content that has been removed, does Discord disclose demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If so, please provide a specific citation to where Discord discloses that information. If not, why not?

No. As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, nor does Discord track the demographic data of users who may view an individual piece of content, regardless of whether the content was deemed violative of our Community Guidelines or the demographics of users who may have viewed the content.

f. Does Discord measure the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If not, why not?

No. As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, regardless of whether the content was deemed violative of our Community Guidelines.

g. Does Discord disclose the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If so, please provide a specific citation to where Discord discloses that information. If not, why not?

No. As explained in the answer to question 1b, Discord does not track the number of times a user views a piece of content on the platform or impressions on content, regardless of whether the content was deemed violative of our Community

Guidelines or the demographic factors of users who may have viewed the content.

3. Discord utilizes an algorithm to recommend or amplify content to users.

Discord is a real-time communications platform that does not rely on algorithmic virality to drive usership. Discord does not have awards, quantification, or down-vote or up-vote features that incentivize users to post certain types of content over others. Discord does not use algorithms to recommend or amplify content to users from servers of which they are not already a member. For communities to which users already belong, Discord may send notifications regarding messages that have received a significant amount of engagement (for example, a message receiving a disproportionately high number of emoji reactions or replies specific to the message, demonstrating substantial engagement with the message, compared to the typical message in the server) and discussions to that user. Before a message can be highlighted to users via a notification, the message must satisfy numerous safety criteria such as whether the message had already been reported to Discord or if the message was shared in a server with adult content text-channels.

a. For content that has been removed, does Discord measure whether and the extent to which the removed content was recommended or amplified by Discord? If not, why not?

As noted above, Discord developed tools to reduce the likelihood that user notifications included content that was later removed due to Community Guidelines violations. Discord does evaluate the effectiveness of its filters to understand how it might improve them over time.

b. For content that has been removed, does Discord disclose whether and the extent to which the removed content was recommended or amplified by Discord? If so, please provide a specific citation to where Discord discloses that information. If not, why not?

Discord does not disclose this information because Discord does not recommend or amplify content algorithmically in the same ways traditional social networks do.

c. For content that has been removed, does Discord measure how many views the removed content received after having been recommended or amplified? If not, why not?

Discord does not measure how many views a single piece of content received because the metric is not relevant to its business. Discord does measure whether a user interacted with a particular notification, but because more than 90 percent of all users are only in servers with 10 or fewer members, "views" are not an

important metric for Discord.

d. For content that has been removed, does Discord disclose the number of views the removed content received after having been amplified or recommended? If so, please provide a specific citation to where Discord discloses that information. If not, why not?

No, for the reasons described above.

4. Does Discord support creating industry-wide transparency requirements to disclose basic safety information, like those included in the *Platform Accountability and Transparency Act*?

Discord is broadly aligned with the goals of the Platform Accountability and Transparency Act. For example, Discord supports greater transparency around safety information, as exemplified by the disclosures that Discord makes in our quarterly Transparency Reports. We also support efforts to promote research into child and teen safety online. However, we think it is important that transparency measures take account of material differences in the designs of online platforms and services. We look forward to working with the Committee as they explore these issues.

Senator Cory Booker
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **Trust and safety teams are a vital component in combatting the spread of CSAM, hate speech, violence, and other violative content on tech platforms. Despite this, tech companies have time and time again disinvested from their trust and safety team, especially during changes in leadership.**

   a. **How has the size of your trust and safety team changed over the past five years? Please provide numbers for each of the past five years.**

   More than 15 percent of our employees work on safety at Discord. By way of context, that means our Safety team is larger than our Marketing team. In addition to the significant investment we have made in our full-time employees working on these issues, we also work closely with external partners to supplement our own work on critical safety issues.

   2019: 22 full-time employees (FTEs), 1 external partner, approximately 10 external partner agents
   2020: 37 FTEs, 1 external partner, approximately 25 external partner agents
   2021: 81 FTEs, 2 external partners, approximately 65 external partner agents
   2022: 79 FTEs, 3 external partners, 120+ external partner agents
   2023: 90 FTEs, 3 external partners, 250+ external partner agents
   2024: 74 FTEs and over 400 additional contract agents including external, virtual Special Operations Center, other support functions

   b. **Do your trust and safety teams make submissions to the National Center for Missing & Exploited Children's CyberTipline, or is that a separate unit?**

   Yes, Discord T&S submits reports to the National Center for Missing and Exploited Children's (NCMEC) CyberTipline when we identify child sexual abuse material or evidence of child sexual exploitation.

   Specifically, this work is the responsibility of our Trust & Safety–Minor Safety and Exploitative Content team. We currently have 17 full-time employees, including three managers.

   c. **If it is a separate unit, how many members are on the team and how have those numbers changed over the past five years. Please provide numbers for each of the past five years.**

   In August 2019 Discord created a dedicated group of six Discord employees to focus on minor safety on the platform—the Minor Safety and Exploitative Content team. Since then, we have grown the team to include 17 dedicated

Discord employees whose sole focus is minor safety and exploitative content.

Additionally, we contract with an external partner that has 134 dedicated agents reviewing a variety of harm types, including non-consensual adult intimate media reports, child sexual abuse material (CSAM) media and grooming reports, and underage user reports.

2. **The National Center for Missing & Exploited Children's CyberTipline plays an integral role in combatting child sexual exploitation. The tipline helps law enforcement investigate potential cases and allows prosecutors to bring justice to victims. While federal law requires your company to report to the CyberTipline any apparent violations of federal laws prohibiting child sexual abuse material of which you are aware, there are many gaps.**

   a. **Is there a standard format your reports to the CyberTipline follow? If so, what is that format?**

      We have a template document that aids our team in matching the relevant information (profile attributes, messages, server and channel attributes, etc.) with the CyberTipline's required fields.

   b. **Does your company proactively report planned or imminent offenses?**

      Yes.

   c. **Does your company proactively report potential offenses involving coercion or enticement of children?**

      Yes.

   d. **Does your company proactively report apparent child sex trafficking?**

      Yes.

Senator Alex Padilla
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **In recent years, more companies in the tech sector are offering tools to enable caregivers to have a dialogue with minors in their care about healthy and safe internet activity. An important element in understanding whether these tools are helpful is understanding whether or not these tools are being adopted.**

   a. **How many minors are on Discord?**

   As of February 2, 2024 there are approximately 2.7 million monthly active users in the U.S. who submitted an age under the age of 18.

   b. **Of these minors, how many of them have caregivers that have adopted your Family Center tool?**

   Discord's Family Center was designed to provide parents and guardians with tools to better understand how their teens use our service. Our Family Center allows parents and guardians to get insights into the communities their teen children have joined, including which users their teens have recently added as friends, which users their teen has messaged or called in DMs or GDMs, and which servers their teen has joined or participated in. The Family Center also allows parents and guardians to get a weekly email with a high-level summary of the time their teen spends on Discord. The Family Center is designed not just to provide parents and guardians with greater insight into how their teen uses Discord, but to also help parents start conversations with their teen about online safety.

   As of February 2, 2024, there are approximately 15,000 parent users currently connected to approximately 15,500 under-18 users utilizing Discord Family Center. Approximately 2.3 in every one thousand users between the ages of 13 and 18 are connected to a parent or guardian through the tool, and early results indicate that nearly 75 percent of parents who utilize the Family Center at least once continue to become regular users of it.

   c. **How are you ensuring that young people and their caregivers are aware of these tools?**

   Discord prominently features the Family Center as part of users' safety settings. We continue to highlight our Family Center tool prominently within our Safety

Center which includes a special resource hub for parents.

Launching the Family Center was a critical investment for Discord in 2023. Discord engaged in a concerted effort to ensure it was advertising to a parent audience, including working with parent influencers, and to highlight the Family Center in parent-targeted news media, such as Parents.com. Discord also employed a mix of paid advertising (video, podcast ads, influencers, search engine marketing) to educate both teens and parents about the tools Discord creates to help teens stay safe on our platform.

Discord also worked closely with the National PTA in the context of its broader partnership to share updates about this feature through their communication channels, including email updates and social media. We also continue to prominently feature Family Center in Discord's Family Center Hub and as a part of Connect Safely's Parent's Guide to Discord.

d. How are you ensuring that these tools are helpful to both minors and their caregivers?

The goal with Discord's Family Center was to design an easy to use, opt-in tool to help create mutually beneficial ways for parents and teens to have conversations about online safety. Using a QR code, parents and teens are able to easily connect their accounts. Parents are both able to view their teen's recent activity in Family Center and will also receive a weekly summary directly to their inbox. In order to help build trust while also respecting teen autonomy, we do share teen message content with parents. Discord continually reviews the utility of its safety tools to ensure that we are able to meet our obligations with regard to child safety.

2. In addition to keeping parents informed about the nature of various internet services, there's a lot more we need to do to inform our young people about unsafe, criminal conduct that is facilitated online. While many companies offer a broad range of "user empowerment" tools, it's helpful for us to understand whether young people even find these tools helpful or are actually adopting them.

a. Discord recently rolled out "Teen Safety Assist" which provides teens safety alerts on first time Direct Messaging senders and a sensitive content filter for teens by default. Have you assessed how these features are impacting your minor users' safety online?

Discord takes a "safety by design" approach to our work—that means a company-wide commitment to building our service safely by assessing risks,

adopting mitigating measures, and evaluating metrics throughout the product development cycle, not waiting until after a product is launched.

Prior to implementing the Teen Safety Assist features, Discord conducted experiments to help us evaluate the efficacy of these tools. Our internal pre-launch review demonstrated that the safety warning delivered via the Teen Safety Assist tool resulted in a significant increase in user-generated reports from direct messages and an increase in users blocking senders that caused the tool to issue a warning to teen users.

Discord continually reviews the efficacy of its safety tools to ensure that we are able to meet our obligations with regard to child safety.

b. How are you ensuring that the burden is not on young people to make adult-level decisions about safety on the services that you operate?

Teens deserve to have a safe, welcoming space to explore their interests, connect with others, and find a place to belong. We're committed to providing a safe, welcoming environment to teens. To do so, we take a multi-faceted and holistic approach to teen safety. This includes:

- an emphasis on education and learning from mistakes over punishment when appropriate with our Warning System;
- safety tools and resources that give users, parents, and educators control to customize their Discord experience for themselves or their teens;
- regularly updating our teen and child safety policies;
- partnerships with industry-leading organizations like the Family Online Safety Institute, Technology Coalition, National PTA, Digital Wellness Lab, INHOPE, and others;
- a trusted reporter process that allows trusted partners to surface content and reports of violations directly to our Safety team; and
- proactive detection and actioning of high-harm content.

We also invest in new ways to detect the exploitation of minors. In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. One of the tools announced as a part of this new initiative is Sensitive Content Filters, which automatically blur potentially sensitive media sent to teens in direct messages (DMs), group direct messages (GDMs), and in servers. Blurring is enabled by default for teen users both in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users. Adult users can

also choose to opt-in to this feature by changing their Privacy & Safety Settings.

Last fall, Discord also launched Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn that leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify a teen user about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

c. Over the last 4 years, how often have you blocked products from launching because they were not safe enough for children, or withdrawn products from the market after receiving feedback on the harms they were causing?

Discord takes a "safety by design" approach to our work—that means a company-wide commitment to building our service safely by assessing risks, adopting mitigating measures, and evaluating metrics throughout the product development cycle, not waiting until after a product is launched. We don't track products we have blocked from launching because we build safety into the design of our products from the start.

3. Existing detection tools for keeping child sexual abuse material from spreading online rely on hashed images of already identified CSAM imagery. There are tools like PhotoDNA and Google's CSAI match tool available for identifying this content. A challenge I hear raised frequently is identifying and removing novel images that have not already been hashed.

a. What would it take to develop better technology to accurately identify and

limit the spread of novel CSAM images?

The technology industry's work to build robust child safety measures and to expeditiously detect, remove, and report child sexual abuse material (CSAM) content would benefit from the ability to train AI and machine learning models designed to detect CSAM on content that platforms have removed but preserved for law enforcement investigations and reporting to the National Center for Missing and Exploited Children (NCMEC). Moreover, Discord and industry peers would be better able to collaborate on innovative approaches to detecting CSAM if companies were able to train detection models on content detected on each other's platforms, as scale is critical to the efficacy of these technologies.

b. Are there interventions from Congress that would facilitate identification of CSAM?

Congress should ensure that current legislative proposals focused on preventing the distribution of CSAM and fighting child exploitation do not unintentionally divert resources away from innovative efforts by online services to stay ahead of bad actors who work hard to avoid detection or make it harder for law enforcement to investigate and prosecute perpetrators of these egregious crimes. Congress could also explicitly exempt work done by companies to develop technology to detect CSAM from the laws prohibiting creation, possession, and distribution of CSAM.

c. Based on your company's experience trying to address online sexual exploitation and abuse of minors, are there areas where Congress could be helpful in tackling this problem?

Congress should provide additional support to NCMEC and should also support and promote the efforts of industry groups and third-party organizations that are dedicated to tackling these issues. For instance, the Tech Coalition is working to strengthen how companies enforce their child safety policies, including through Lantern, a cross-platform signal sharing program.

Congress should also assist federal and state law enforcement units dedicated to child safety issues. These units—which play an essential role in tackling these challenges—are under-funded and under-resourced.

4. AI models are making it easier to develop synthetic CSAM. These are either altered images of real people, or wholly synthetic individuals. Policymakers are grappling

with what this will mean for law enforcement efforts to hold perpetrators accountable and identify children who are being harmed. In addition to processing a higher volume of Cybertips, investigators will have the added challenge of determining whether the victim in the scenario is in fact a real person. And cases are already being reported where AI generative technologies are being employed to facilitate the grooming and sextortion of minor victims.

a. **What are you doing to identify and remove AI-generated CSAM on your services?**

In addition to our use of industry standard tools to detect and remove known CSAM, Discord invests in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online.

b. **Do you flag for NCMEC if you perceive the CSAM to be AI-generated?**

Yes, we report AI-generated CSAM to NCMEC. Additionally we take action against users (and in severe cases may report users to NCMEC) who attempt to prompt AI bots on our service to create CSAM through requests that sexualize children textually.

c. **How prevalent is this kind of content?**

Discord's decision to invest in proactively detecting and removing servers before they are reported to us, especially for high-harm categories, has paid dividends in user safety. In 2023 alone, Discord scanned more than 47 billion images, or roughly 130 million images every single day. Images generated through AI models are monitored via the same visual detection system as all other messages and images on the platform. As a result, we scan the attachments stored on Discord for CSAM via hash matching. For instances where the user submits images to modify with AI, the input images are scanned for CSAM as well.

Additionally, images generated through AI models are monitored via the new visual detection tools used to detect unknown and AI-generated CSAM, which Discord built in collaboration with industry peers.

During the fourth quarter of 2023, Discord proactively removed servers for child safety 96 percent of the time; servers in which CSAM was posted were removed proactively 97 percent of the time. Overall, 94 percent of servers removed for policy violations across categories during this period were removed proactively. We continue to invest in our ability to proactively detect and remove servers before they are reported to us, especially for high-harm categories.

At the account level, in the fourth quarter of 2023 Discord reported 55,955 accounts to NCMEC through our use of PhotoDNA and hashing systems such as our visual safety platform, PDQ, and CLIP. 55,638 of those reports were media (images or videos). Additionally, 317 high-harm grooming or endangerment reports were delivered to NCMEC. Discord disabled 116,219 accounts and removed 29,128 servers for child safety during this same period.

d. How do you anticipate the rise of AI-generated CSAM will impact NCMEC's ability to process and refer Cybertips to law enforcement?

The proliferation of AI-generated CSAM poses challenges for industry, law enforcement, and civil society. In our work to combat the spread of CSAM online we value our partnership with NCMEC and support efforts to significantly increase NCMEC's funding and resources to address both the increase in Cybertips and the growing complexity of their work.

e. Recently, A.I.–generated explicit images of a major pop superstar were distributed widely online without her consent. That story drew attention to a growing problem over the last year facilitated by AI tools: the generation of deepfake, nonconsensual, sexually explicit imagery of everyday people, including our young people. Will you commit to reporting on the prevalence of this new problem and the steps your company is taking to address this horrendous abuse?

Discord's Community Guidelines prohibit the posting of sexually explicit or sexually suggestive content of other people without the subject's knowledge and consent. This includes the non-consensual distribution of intimate media that was created either with or without an individual's consent.

Technological challenges currently frustrate the ability to distinguish accurately between actual and AI-generated non-consensual sexual explicit images, and thus to collect data and report on the prevalence of violative content that is AI-generated.

f. Are there technical or legal barriers that your company has identified preventing thorough redteaming of AI models to ensure they do not generate CSAM?

While there are no strict legal barriers, there is no explicit exemption to the laws prohibiting the creation, possession, and distribution of CSAM for the work conducted by companies engaging in redteaming of AI models.

5. How companies choose to allocate their resources illustrates their true priorities.

a. What percentage of your company's budget is dedicated to addressing child safety on your platform?

Discord's budget is not organized in a manner which allows it to determine funding dedicated to solely addressing online child safety-related components, but substantial resources from Trust & Safety, Engineer, Legal, Policy, and other teams work towards addressing this issue.

b. What process or assessment of risk on the platform informed that figure?

Discord annually undertakes a holistic budgeting process in order to maximize the impact of its resourcing across its company priorities, including Trust & Safety.

c. How many layers of leadership separate your trust and safety leaders from you?

None: John Redgrave is Discord's Vice President in charge of Trust & Safety and leads a cross-functional Safety team. John reports to Clint Smith, Discord's Chief Legal Officer, and Clint is responsible for our Law, Policy, and Safety functions and reports directly to me.

6. The companies represented at the hearing have the money and resources to hire teams of Trust & Safety professionals and build bespoke tools to aid with content moderation and integrity work as well as the detection of content like CSAM on their services. This is not necessarily the case for the rest of the tech sector. These

are industry-wide problems and will demand industry-wide professionalization and work.

a. **What is Discord currently doing to support access to open-source trust & safety tools for the broader tech ecosystem?**

Discord is a funder and supporter of an open source tooling hub for technology that promotes online safety, an initiative that aims to make online safety tools broadly available to all platforms. Information on this can be found in this recent announcement from Columbia University. As part of the Columbia initiative, Clint Smith, Discord Chief Legal Officer who is responsible for Law, Policy & Safety, explained that: "Technology that improves online safety should never be one company's competitive advantage, but should be shared for the common good. That's why Discord strongly endorses taking a more structured approach to the development and distribution of open source Safety technology, and we're proud to join with Google and leading philanthropic organizations to support [Columbia's] work to advance this vision—for the benefit of society and the industry."

A recent example of success in this area is Discord's development of innovative technology to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source and available through the Tech Coalition, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online. By leveraging these tools, Discord's model had an 87 percent increase in recall, i.e. proactively-detected confirmed positive CSAM.

7. **One necessary element of keeping our kids safe is preventing harms in the first place. The National Center for Missing and Exploited Children partnered with the White House, the Department of Justice, and the Department of Homeland Security to create "The Safety Pledge" initiative to combat online child exploitation in September 2020. I understand more government backed public awareness campaigns are being developed.**

a. **Are you partnering with the federal government to distribute health and**

safety resources to young people?

No.

b. What are you proactively doing to educate the minors that use your services about online health and safety?

Discord's work with parents and parent-led organizations is vital to our child safety efforts. Since 2022, Discord has been a sponsor of the National PTA. Discord collaborates with National PTA on digital safety resources and events designed to elevate youth voices and foster shared understanding among teens, caregivers and educators. The PTA Connected: Build Up and Belong program, sponsored by Discord, facilitates technology discussions between teens and their parents and caregivers about online safety topics. In 2023, 25 local PTAs were selected to receive $1,250 each, sponsored by Discord, to host a PTA Connected: Build Up and Belong program. Discord also sponsored grants for this program in 2022.

Discord is also a member of the Family Online Safety Institute (FOSI). FOSI develops resources to provide parents and guardians with the tools they need to confidently navigate the online world with their families.

Discord also supports ConnectSafely, a non-profit organization that educates people about online safety, privacy, security, and digital wellness. ConnectSafely has developed research-based safety tips, as well as in-depth guides and "quick-guides" for parents, educators, youth, and policymakers. Discord worked with ConnectSafely to develop a Parent's Guide to Discord.

In recognition of the unique safety challenges that accompany teens' use of the internet, Discord has also created special features designed to give parents and teens even more control over their experience on Discord. The Family Center was designed to provide parents and guardians with tools to better understand how their teens use our service. Our Family Center allows parents to get insights into the communities their teen children have joined, including which users their teens have recently added as friends, which users their teen has messaged or called in DMs or GDMs, and which servers their teen has joined or participated in. The Family Center also allows parents to get a weekly email with a high-level summary of the time their teen spends on Discord. The Family Center is designed not just to provide parents and guardians with greater insight into how their teen uses Discord, but to also help parents start conversations with their teen about

online safety.

As part of our Safer Internet Day efforts in 2023 we partnered with NoFiltr, the youth-targeted online safety program by Thorn, to develop online safety resources for young people. These resources included an interactive quiz, online fortune teller, and a social media campaign aimed at supporting healthy digital habits and online safety.

Finally, we invest heavily in advanced tooling and education so parents know how our service works and understand the controls that can contribute to creating a positive and safe experience on Discord for their children. Discord also continues to innovate our approach to child safety. As described above, our recently announced Teen Safety Assist initiative contains new tools to keep teens safe on Discord: Sensitive Content Filters and Safety Alerts for Senders are already in place and working to make Discord a safer place for teens; Safety Alerts in Chat will be rolled out in the coming weeks.

In addition, we strive to understand what is top of mind for teens and to build safety tools and resources that resonate with their lived experiences. In 2023, as a part of Safer Internet Day, we released a fortune teller with questions and icebreaker prompts to help jump start a conversation about better digital health and safer online practices.

8. **Sextortion has become increasingly prevalent. Offenders may use grooming techniques or basic trickery to manipulate victims into providing nude or partially nude images of themselves, which are then used to coerce victims into sending more graphic images and videos or pay a ransom. These criminals often threaten to post the images or sensitive images publicly or send them to the victim's friends and family if the child does not comply. From May 2022 to October 2022, U.S. law enforcement and NCMEC witnessed an alarming increase in Cybertips and reports where minors have been sextorted for money. Many young boys, including in California, have committed suicide out of desperation, leaving their loved ones devastated.**

    a. **How is your company responding to the growing threat of financial sextortion?**

    As with our efforts to prevent and disrupt the dissemination of CSAM, Discord's approach to child safety more broadly leverages a mix of both proactive detection and reactive measures, supported by a range of human-powered and technical

solutions. Examples relevant to the issue of sextortion include Discord's Safety Alerts for teen users, which are designed to detect unwanted interactions via text message and better equip teens to safely navigate such situations.

In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. Among the tools announced as a part of this new initiative are Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, which leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

Additionally, Discord has also invested in staffing a dedicated team housed under our Trust & Safety Minor Exploitative Content team whose sole focus is addressing high-harm sextortion activity on the platform.

b. What methods are in place to detect and disrupt this type of abuse in real time?

Please see the answer to the preceding question.

c. What kind of user education and awareness are you engaged in?

As described in detail in response to question 7(b), Discord's user education and awareness efforts include our partnerships with and support of parent- and guardian-led organizations, our direct outreach to parents and guardians to explain

how Discord works and the parental controls available to them, and our work to build new safety features for our teen users, such as those announced as a part of our Teen Safety Assist initiative.

d. Are you aware of a higher prevalence of sexual extortion or abuse against certain demographics among young users? If not, will you commit to studying this issue and making that kind of information available to improve public education and protection measures?

Discord is not aware of a prevalence of sexual extortion or abuse targeting certain demographics among young users. Discord continually reviews the efficacy and utility of its safety tools to ensure that the tools we deploy take proper measure of the online safety threats currently facing teens. Discord is committed to continued collaboration with industry peers to best confront these challenges.

9. Young people need to be at the center of regulatory discussions, and they need to be at the table as products and services they use are designed.

a. Are you engaging young adults and youth in your conversations and policies around Trust and Safety on the platform?

We strive to understand what is top of mind for teens and to build safety tools and resources that resonate with their lived experiences. We have also recently launched a Teen Charter project. As part of this effort, we are working with experts and youth councils and running our own focus groups to develop an aspirational charter that will help us both understand how teens want to feel on Discord and informs the ways we build new products and features to keep them safe on the platform. For example, Discord engaged with the NoFiltr Youth Innovation Council on the Teen Charter project as well as our broader teen safety efforts to better understand teen attitudes and concerns on topics related to online safety. In addition, Discord seeks to collaborate with groups, such as Digital Wellness Lab, that prioritize youth participation.

b. How do you proactively keep up to speed with the most pressing issues facing young people online?

Discord's Teen Charter project is designed to help us stay up to speed on the issues young people face online. We also follow the latest research in this space and engage with experts such as Digital Wellness Lab and Tech Coalition to ensure we are constantly updating our approach. We have hired a teen safety

policy manager and mental health policy manager to further aid in work.

10. **For many children, an open dialogue about their internet habits is a best practice, and healthy. But not every child has a parent or a caregiver that is looking out for their best interest. For many kids who are abused, a caregiver or parent is their abuser. Additionally, for many young people, their parents' knowledge of their sexual orientation or their interest in exploring it, fundamentally puts them in jeopardy. Solving for these different needs across our young people at the scale of social media and internet applications is really vital.**

    a. **How have you designed your parental tools with this dynamic in mind?**

    The Family Center was designed to provide parents and guardians with tools to better understand how their teens use our service. Our Family Center allows parents and guardians to get insights into the communities their teen children have joined, including which users their teens have recently added as friends, which users their teen has messaged or called in DMs or GDMs, and which servers their teen has joined or participated in. The Family Center also allows parents to get a weekly email with a high-level summary of the time their teen spends on Discord. The Family Center is designed not just to provide parents and guardians with greater insight into how their teen uses Discord, but to also help parents start conversations with their teen about online safety.

Senator Laphonza Butler
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **Family and parental control tools: I was glad to hear that you have spent time talking with parents and what their families need from your products. I was also glad to hear your companies have a Family Center, or other similar tools, to give parents more insight and control over how their children are using your platforms and apps.**

   a. **How do you advertise this feature to parents?**

   Discord prominently features the Family Center as part of users' safety settings. We continue to highlight our Family Center tool prominently within our Safety Center which includes a special resource hub for parents.

   Launching the Family Center was a critical investment for Discord in 2023. Discord engaged in a concerted effort to ensure it was advertising to a parent audience, including working with parent influencers, and to highlight the Family Center in parent-targeted news media, such as Parents.com. Discord also employed a mix of paid advertising (video, podcast ads, influencers, search engine marketing) to educate both teens and parents about the tools Discord creates to help teens stay safe on our platform.

   Discord also worked closely with the National PTA in the context of its broader partnership to share updates about this feature through their communication channels, including email updates and social media. We also continue to prominently feature Family Center in Discord's Family Center Hub and as a part of Connect Safely's Parent's Guide to Discord.

   b. **Can you share data on how many Family Center/parental tools users there are in proportion to total minors on your platforms and products?**

   As of February 2, 2024, there are approximately 15,000 parent users currently connected to approximately 15,500 under-18 users utilizing Discord Family Center. Approximately 2.3 in every one thousand users between the ages of 13 and 18 are connected to a parent or guardian through the tool, and early results indicate that nearly 75 percent of parents who utilize the Family Center at least once continue to become regular users of it.

Senator Chuck Grassley
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

Please answer each question to the fullest possible extent. If your platform is unable to answer a particular question or does not have requested data, explain why. Each question refers to your company in addition to any corporate affiliates, including parent and subsidiary companies.

1. **Current law requires that a provider of a report of suspected CSAM to the National Center for Missing and Exploited Children's (NCMEC) CyberTipline preserve "any visual depictions, data, or other digital files that are reasonably accessible and may provide context or additional information about the reported material or person" for a minimum of 90 days. 18 U.S.C. 2258A(h)(1-2). The recent explosion of suspected abuse has presented unprecedented challenges for law enforcement to follow up on leads before companies discard or delete essential data and information. There is nothing preventing tech companies from preserving relevant material beyond the statutorily-mandated 90-day period.**

    a. **How long does Discord voluntarily preserve and retain data contained in and related to its reports to the CyberTipline?**

    Discord voluntarily preserves content reported to NCMEC's CyberTipline for 2 years.

    b. **The massive influx of reports to the CyberTipline naturally results in law enforcement entities having to conduct and finish investigations beyond 90 days of an initial report to the CyberTipline. Retaining relevant information for longer periods could significantly advance law enforcement's ability to thoroughly investigate leads. If Discord only preserves and retains this information for the minimum 90-day period, why does it do so when preserving this data longer could significantly enhance and prolong law enforcement's ability to investigate and prosecute child predators?**

    Discord voluntarily preserves content reported to NCMEC's CyberTipline for 2 years.

    c. **Please confirm if Discord stores and retains the following information relating to reports to the CyberTipline:**

i. **IP addresses**

Yes.

ii. **Screen Names**

Yes.

iii. **User Profiles**

Yes.

iv. **Associated Screennames (by IP address and associated emails)**

Yes.

v. **Email addresses**

Yes, where available.

vi. **Geolocation data**

Discord does not collect geolocation data.

d. If Discord does not retain or store any of the above types of information in question (c), please explain why.

Discord does not collect geolocation data, and thus does not retain geolocation data.

e. Please list any other information Discord retains and preserves for law enforcement purposes not listed above in question (c).

When an account is preserved for law enforcement purposes, all relevant content associated with the account is captured at the time of the preservation and retained.

f. Does Discord flag screennames and associated email addresses to suspected accounts that violate Discord's terms of service?

Under certain circumstances Discord will proactively preserve and disclose information to law enforcement, including screennames and associated email addresses. These circumstances are generally for emergency situations when we possess a good faith belief that there is an imminent risk of serious physical injury or death.

2. **How does Discord prioritize urgent requests for information from law enforcement and what is Discord's response time to urgent requests?**

   Discord reviews and responds to all legal requests from law enforcement in a timely manner. We comply with valid legal process before any indicated due date and communicate with law enforcement on the status of their request. For emergency disclosure requests, Discord strives to respond within 15 minutes and have a full resolution within an hour of receipt.

3. **What is Discord's average response time to service of legal process from law enforcement for CSAM-related information?**

   Discord reviews and responds to all legal requests from law enforcement in a timely manner. We comply with valid legal process before any indicated due date regardless of the related crime. Notably, law enforcement requests do not necessarily specify whether the information sought is related to a child sexual abuse material (CSAM) investigation. Discord is unable to disaggregate CSAM-related process from other legal process from law enforcement and as such cannot provide a response to this portion of the question.

4. **In 2023, the tech industry as a whole slashed more than 260,000 jobs. And in the first four weeks of this year, another 25,000 jobs were cut.**

   a. **For each year, between 2018 and 2023, how many U.S. based employees did you have at Discord?**

      2018: 151 U.S. based employees
      2019: 217 U.S. based employees
      2020: 308 U.S. based employees
      2021: 669 U.S. based employees
      2022: 1,008 U.S. based employees
      2023: 1,126 U.S. based employees

      i. **Of these employees, how many were sponsored on H-1B visas?**

2018: 2 employees were newly sponsored on H-1B visas
2019: 3 employees were newly sponsored on H-1B visas
2020: 3 employees were newly sponsored on H-1B visas
2021: 12 employees were newly sponsored on H-1B visas
2022: 26 employees were newly sponsored on H-1B visas
2023: 9 employees were newly sponsored on H-1B visas

ii. For each year, between 2018 and 2023, how many H1-B visa applications did Discord submit?

Below includes applications submitted, including extensions and amendments to applications. This also includes applications where the individual ultimately did not join the company.

2018: 5
2019: 3
2020: 4
2021: 15
2022: 30
2023: 17

b. For each year, between 2018 and 2023, how many employees based outside the U.S. did you have at Discord?

2018: 0
2019: 0
2020: 0
2021: 6 employees outside the U.S.
2022: 35 employees outside the U.S.
2023: 41 employees outside the U.S.

i. Of these employees, how many were based in China?

0.

c. For each year, between 2018 and 2023, how many employees in total did Discord terminate, fire, or lay off?

2018: 6 employees were involuntarily separated
2019: 9 employees were involuntarily separated

2020: 9 employees were involuntarily separated
2021: 10 employees were involuntarily separated
2022: 38 employees were involuntarily separated
2023: 95 employees were involuntarily separated

i. **Of these employees, how many were based in the United States?**

2018: 6 U.S. employees
2019: 9 U.S. employees
2020: 9 U.S. employees
2021: 10 U.S. employees
2022: 37 U.S. employees
2023: 90 U.S. employees

ii. **Did Discord fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?**

No.

iii. **Were any duties and/or functions previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?**

Discord does not have a system by which individual duties and functions of previous employees are tracked upon their departure and thus cannot answer this question.

d. **For each year, between 2018 and 2023, how many employees performing work related to child safety did Discord terminate, fire, or lay off?**

2018: 0
2019: 0
2020: 2 employees were involuntarily separated
2021: 0
2022: 0
2023: 1 employee was involuntarily separated

i. **Of these employees, how many were based in the United States?**

2018: 0

2019: 0
2020: 2
2021: 0
2022: 0
2023: 1

ii. **Did Discord fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?**

No.

iii. **Were any duties and/or functions (specifically relating to child safety) previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?**

Discord does not have a system by which individual duties and functions of previous employees are tracked upon their departure and thus cannot answer this question.

iv. **How have layoffs impacted Discord's ability to protect children on its platforms?**

Notwithstanding Discord's recent reduction in force, the percentage of employees working on safety at Discord remains unchanged. More than 15 percent of our employees continue to work on safety at Discord. By way of context, that means our Safety team is larger than our Marketing team. In addition to the significant investment we have made in our full-time employees working on these issues, we also work closely with external partners to supplement our own work on critical safety issues.

v. **Does Discord have any plans to increase staff responsible for child safety operations or otherwise optimize its child safety operations?**

Minor safety is a top priority for Discord's Trust & Safety team, and we have shifted resources and internal subject matter experts to focus on keeping teens safe on our platform. Minor Safety is the largest single team inside of Discord's broader Trust & Safety team, with over 20 percent of our total full-time employees and 40 percent of our external partners working in this area.

5. **On January 30, 2024, the Tech Transparency Project (TTP) published an article on their website called, "Meta Approves Harmful Teen Ads with Images from its Own AI Tool". In summary, TTP, using Meta's "Imagine with Meta AI" tool generated inappropriate images such as young people at a pill party or other vaping. These images with text were submitted to Facebook as advertisements targeting users between ages 13-17 in the United States. TTP reported that Facebook approved the advertisement, despite it violating its own policies, in less than five minutes to run on the following platforms: Facebook, Instagram, Messenger, and Meta Quest. Meta. Over the course of a week, TTP submitted the advertisements with the same end result: Facebook approving them. TTP reported that they canceled these advertisements before their scheduled publication, but it illustrated the repeated failures of Facebook to properly moderate content. This is just one example of what other non-government organizations and others have uncovered across social media platforms.**

    a. **How often a month do Discord employees conduct quality checks on Discord's policies and safeguards for child accounts?**

    Discord conducts quality assurance on actions taken by employees with respect to user violations (i.e., we have quality control measures in place to ensure our decisions are correct). Our Policy team generally reviews and updates our Community Guidelines once a year, consistently updates platform guidelines to meet industry standards for child safety, including updating policies for violative generative AI images, and we publish those updates in Discord blog posts.

    b. **In which departments, components, or units of the company does Discord have staff dedicated to performing this type of work?**

    Our T&S team has quality assurance (QA) standards set by a dedicated T&S QA team. We have QA specialists that verify actions taken by full-time employees (FTEs), an Appeals team to verify correct actions taken through the appeals process, and a QA team dedicated to examining the quality of actions taken by our external partners. The external partners have internal QA teams (separate from Discord teams) that ensure actions taken are up to standards we have set across T&S.

    The information provided above does not include employees on cross-functional teams that contribute to trust and safety, including engineering, marketing, policy, legal support, and more who work on child safety and safety more broadly.

c. How many employees make up these departments, components, or units?

Discord has 2 full-time employees dedicated to quality assurance checks of our external partners. More than 20 percent of our Trust & Safety full-time employees are involved in QA to some extent. In total, we have 74 FTEs in T&S at Discord.

The information provided above does not include employees on cross-functional teams that contribute to trust and safety, including engineering, marketing, policy, legal support, and more who work on child safety and safety more broadly.

d. If a violation is found, what action is taken, and how quickly is action taken?

Discord continually reviews the efficacy of its own internal processes and procedures to ensure that we are able to meet our obligations with regard to child safety. Where the business determines that our work has not followed established internal processes, we work quickly to take corrective action.

6. Social media companies claim they are investing in company components dedicated to safety, and that their platforms are safe for children. However, children continue to be exploited daily across these platforms.

a. What have Discord's revenue and profit figures been for the last three years (2021-2023)? Please provide figures broken out per year. Do not provide percentages.

Discord is a private company and does not publish or share revenue figures publicly. We'd be happy to provide these figures to your office confidentially.

b. How much has Discord spent in advertising for the last three years (2021-2023), broken out per year?

2021: $31.0M
2022: $5.1M
2023: $2.3M

c. How much of Discord's resources spent on advertising has been devoted to advertising Discord's safety initiatives and efforts for the last three years (2021-2023), broken out per year?

Prior to 2023 Discord did not disaggregate resources spent on advertising based

on the initiatives they applied to and thus cannot provide numbers prior to 2023. In 2023, Discord spent $1.7M on advertising devoted to Discord's safety initiatives and efforts.

d. To get an understanding of how your company has invested and plans to invest in its components dedicated to child safety functions, what are the annual budgets for Discord's child safety-related components for the last three years (2021-2023)?

Discord does not track its budget allocation in this way and cannot provide historic budget allocations related only to child safety functions. More than 15 percent of our employees continue to work on safety at Discord. By way of context, that means our Safety team is larger than our Marketing team. In addition to the significant investment we have made in our full-time employees working on these issues, we also work closely with external partners to supplement our own work on critical safety issues.

e. What is the current anticipated (2024) budget for Discord's child safety-related components?

Discord's budget is not organized in a manner which allows it to determine funding dedicated to solely addressing online child safety-related components, but substantial resources from Trust & Safety, Engineer, Legal, Policy, and other teams work towards addressing this issue.

f. Provide the number of staff employed in Discord's child safety-related components for the last three years (2021-2023).

Discord does not track staff allocation in a way that allows us to answer this question at this level of granularity. More than 15 percent of our employees continue to work on safety at Discord. By way of context, that means our Safety team is larger than our Marketing team. In addition to the significant investment we have made in our full-time employees working on these issues, we also work closely with external partners to supplement our own work on critical safety issues.

g. How much is that compared to Discord's other components for the same period? (Please provide a breakout per year. Do not provide percentages.)

Discord does not track staff allocation in a way that allows us to answer this

question at this level of granularity. More than 15 percent of our employees continue to work on safety at Discord. By way of context, that means our Safety team is larger than our Marketing team. In addition to the significant investment we have made in our full-time employees working on these issues, we also work closely with external partners to supplement our own work on critical safety issues.

h. How many staff are currently employed in Discord's child safety-related components?

18 FTEs—nearly 20 percent of our total T&S employees—are dedicated to our Minor Safety efforts. Nearly 40 percent, or approximately 158 external partner agents, of our external partners are directly involved with child safety reports. These numbers do not include other employees and individuals who contribute to Discord's child safety-related components more broadly, including product managers, engineers, data scientists, and other cross-functional roles that work on child safety either full- or part-time.

i. What are the roles, responsibilities, and functions of Discord's child safety-related components?

Discord has subject matter experts dedicated to proactive investigations, responding to user reports, handling escalations from our external partners, outreach to industry peers and partners, collaborating with law enforcement and government notices, building industry standard quality assurance requirements across FTEs and external partner agents, building new technologies in collaboration with our engineering and machine learning teams, and informing and evolving our products and policies in consultation with users, their parents and guardians, and industry experts.

Beyond the individuals dedicated to the above components, many more employees also play an important role in Discord's child safety work. Product and engineering teams develop tools like Teen Safety Assist and Family Center to help keep teens safe, Data Science monitors and assesses the effectiveness of our features and efforts, Policy maintains Discord's policies regarding child safety to ensure alignment with current best practices, and Legal evaluates Discord's legal obligations around child safety.

j. Are any other components responsible for the monitoring of CSAM on Discord's platform(s)?

In 2023, Discord T&S built a virtual security operations center (vSOC) to help monitor the platform for threats, risks, and violations (including CSAM). We work with various partners and vendors who help provide additional signal for potential violations on the platform. Our engineering and machine learning teams are focused on building new and evolving existing technologies for detection and prevention of CSAM on Discord.

k. What, if any, third parties does Discord employ or contract with to address CSAM material on its platforms?

Discord works with a leading service provider who specializes in content moderation. Specifically, we worked with them to create a dedicated Minor Safety and Exploitative Content team to specialize and address child safety issues such as CSAM and grooming on the platform.

i. What are the roles and responsibilities of these third parties?

Discord's service provider's dedicated Minor Safety and Exploitative Content team works with us to analyze, investigate, and enforce outcomes of all child safety-related reports.

This includes:

- Reviewing reported media content (images and videos) to determine whether the content contains CSAM to either report to NCMEC or to escalate to an internal team member for a full investigation on real-world harm;
- Reviewing reported conversations between users or groups of users to determine whether the content contains a (present or not present) child being groomed, sexualized, extorted/threatened, or self-endangered to either perform enforcement actions or escalate to an internal team member for a full investigation on real-world harm;
- Reviewing reported conversations between users or groups of users to determine whether a child has previously or will be groomed, sexualized, extorted/threatened, or self-endangered by an adult user to escalate to an internal team member for a full investigation on real-world harm.

ii. What is the breakdown of cost per third party over the last three

years (2021-2023)?

Discord contractual costs during the years in question are:

2021: $0
2022: $217,600
2023: $1.4M

7. **Of all reports sent by Discord to the National Center for Missing and Exploited Children, how many reports were self-generated from victim users for the last three years (2021-2023)? Please provide the actual number of self-generated reports in addition to the total number of reports (including those that were not self-generated). In addition, please provide a break-down of the self-reporters by age.**

   In 2023, Discord added a feature to its internal reporting system that allowed us to track CSAM submitted to NCMEC that our Trust & Safety team determined to be self-generated CSAM (SG-CSAM). Prior to 2023, our systems did not track these reports as a distinct subset, and so we are unable to share pre-2023 SG-CSAM data here.

   - SG-CSAM Reported to NCMEC (starting in January 2023): 1,519 reports
   - All CSAM Reported to NCMEC (2021-2023): 211,426

8. **What is Discord's policy or protocol with respect to law enforcement accessing user data and subsequent notification to users of law enforcement accessing their data?**

   As an electronic communication service provider, Discord strictly complies with the provisions of the Electronic Communications Privacy Act (ECPA), and discloses information to law enforcement in accordance with the provisions of the statute. With regard to user notice, Discord would not provide a user with notice when we receive an order prohibiting us from doing so. Moreover, if the legal process pertains to child sexual exploitation or the danger of death or serious bodily injury, Discord does not provide notice to help avoid a harmful or dangerous outcome resulting from the notice.

   Discord does not notify users when it receives a 90-day preservation order under the provisions of existing law. See 18 USC 2703(f). Thus, after sending a preservation order to Discord, law enforcement always has 90 days to secure proper legal process, including a non-disclosure order, while relevant materials are preserved.

   Additionally, if Discord receives legal process without a prior preservation order and

without a non-disclosure order, even in cases where Discord intends to provide notice to a user, we would first alert law enforcement of this plan and allow law enforcement to withdraw the subpoena or delay its enforcement until a non-disclosure order can be obtained. This policy effectively ensures that users are not notified of legal process in circumstances where law enforcement wants to prevent such notification. More information on Discord's practices can be found in this article: How Discord Works With Law Enforcement.

a. **Do certain crimes such as drug trafficking or child exploitation affect Discord's decision to notify a user whose data is accessed by law enforcement?**

   Yes. Discord will not notify users in investigations involving child sexual exploitation or the danger of death or serious bodily injury. Discord takes seriously its responsibility to balance the privacy rights of its users and their ability to properly challenge legal process with the legitimate investigatory needs of law enforcement and the safety of the public.

b. **Do certain requests such as a subpoena or search warrant affect Discord's notification protocol? If so, what are they?**

   Discord does not provide notice if we are prohibited by law, such as through a non-disclosure order, or in cases involving exceptional circumstances such as child sexual exploitation investigations or threat to life emergencies. The more detail provided to us, the more easily we can apply these exceptions.

c. **If Discord does notify users of law enforcement accessing their data, why does Discord find this necessary?**

   Discord balances users' right to challenge legal process with the legitimate investigatory needs of law enforcement and the safety of the public. User notification is not only a best practice in the provider community, but also important to allow third-parties to better protect their confidential information when appropriate to do so. As noted above, Discord does not provide notice if we are prohibited by law, such as through a non-disclosure order, or in cases involving exceptional circumstances such as child sexual exploitation investigations or threat to life emergencies.

9. **The National Center for Missing and Exploited Children has indicated that reports from social media companies tend to lack actionable information causing law**

enforcement to be burdened with incomplete information. How comprehensive are Discord's reports to NCMEC? What challenges is Discord experiencing on the collection of user data and other information to include in its reports to NCMEC? What actions is Discord taking to make its reports more comprehensive and useful to law enforcement?

Discord's reports to NCMEC via their application programming interface contain all fields required by NCMEC for their reporting.

Discord's manual reporting (which is often related to potential grooming) process is designed to provide NCMEC an actionable summary of the context giving rise to the report. Agents review archives for both the reported user and potential child victim, investigate and report all relevant discussions between those parties, and report on anywhere we observe age statements from the parties involved in the exchange and any continuing sexual discussions after the reported user was made aware the individual they were engaging with was likely a minor.

Additionally, we review feedback from law enforcement officers that voluntarily disclose the disposition of cases stemming from CyberTip reports filed with NCMEC by Discord. These are emailed to Discord at a monthly cadence and we review reports to ensure that all necessary information for law enforcement was disclosed in the original report.

Senator Mike Lee
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **Currently, Discord restricts certain content from accounts that belong to minors. However, the only age verification measure that Discord undertakes to ascertain the age of its users is asking new users to enter their birthdate when they open an account. How do you prevent minors from creating an account as an adult by lying about their age?**

   When an individual attempts to register a new account, Discord requires the individual to enter their date of birth through a neutral age gate. If the individual provides a date of birth indicating that they are under 13 years old, Discord does not permit the individual to register an account and instead presents a notice screen informing the individual that they are "unable to register." If the individual selects "Back to Login," they are shown the typical registration landing screen and if they select "Register," they are once again shown the "unable to register" notice without being provided an opportunity to make up a new birthdate or attempt to re-register. Discord provides additional information to the public about why the service asks for individuals' date of birth in a published Trust & Safety article.

   Discord also takes steps to confirm that registered users are 18 years of age or older when they attempt to join an "age-restricted" server or channel. As outlined in the Community Guidelines, users must be age 18 or older to participate in communities containing adult content on Discord and servers and channels organized around adult content or age-restricted themes (e.g., violent content) must be labeled as age-restricted. When a user first attempts to join an age-restricted server or channel, Discord implements an extra layer of protection, through an age-restricted warning or gate.

   Discord also investigates and verifies a user's age in situations where either (i) the user, within the platform, admits that they are underage or (ii) Discord receives a credible report, supported by clear and convincing evidence, that the user is underage. For instance, Discord sometimes receives communications allegedly from a parent or guardian asserting that their underage child has a Discord account. Discord's Trust & Safety team takes these reports seriously and follows processes for investigating such reports and, where appropriate, verifying the user's age and disabling the user's account if we determine that the user is under 13.

   Finally, Discord undertakes efforts to identify a user's age when a user whose account has been banned for being underage submits an appeal. Discord employs a rigorous verification process for investigating and arbitrating such appeals with an emphasis on giving the user (or their parent) an opportunity to definitively prove that the user is age 13 or older, while leaving properly executed bans in place.

2. **Digital sextortion is a growing epidemic on all social media platforms. Several victims of sextortion were approached by predators on your platform. You have increased the number of employees tasked with child safety, and you automatically report suspected grooming to NCMEC. How do you ensure that these predators cease contacting minors? How do you catch these situations before a predator attempts to get the minor to move to another platform?**

   As with our efforts to prevent and disrupt the dissemination of child sexual abuse material (CSAM), Discord's approach to child safety more broadly leverages a mix of both proactive detection and reactive measures, supported by a range of human-powered and technical solutions. Examples relevant to the issue of sextortion include Discord's Safety Alerts for teen users, which are designed to detect unwanted interactions via text message and better equip teens to safely navigate such situations.

   In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. Among the tools announced as a part of this new initiative are Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

   On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, which leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

   Additionally, Discord has also invested in staffing a dedicated team housed under our Trust & Safety Minor Exploitative Content team whose sole focus is addressing high-harm sextortion activity on the platform.

3. **Discord automatically blurs sexual content for minor's accounts, but an independent investigation led by NCOSE found that an image containing sexually explicit material sent from an adult account to a fake minor account was not blurred. What are you doing to ensure these instances do not continue to occur?**

Discord is continuously working to improve and refine our detection capabilities and response strategies in order to further reduce our false negative rate on images such as the one described in the question. We are not aware of the specific timeframe of the NCOSE report, but their analysis may have predated the November 2023 launch of the explicit content filter technology in our Teen Safety Assist feature.

4. **When an adult account sends explicit sexual content to a minor's account, is that content permitted? Is the adult account reported to NCMEC or other law enforcement? Can a blurred image still be received by the minor?**

   This content is not permitted under Discord's Terms of Service or Community Guidelines. All age-restricted content must be labeled and is only permitted to be shared in age-restricted areas that exclude minor users.

   We report all instances where we observe: age statements from the suspected adult (reported user) and suspected child (child victim), which are followed by inappropriate sexual conduct with a child to the National Center for Missing and Exploited Children (NCMEC) via their CyberTipline as potential grooming instances.

5. **How does Discord inform parents when a child is exposed to sexual material? How does Discord inform parents when their child is the target of grooming?**

   In recognition of the unique safety challenges that accompany teens' use of the Internet, Discord has created special features designed to give parents and teens even more control over their experience on Discord. Discord does not always possess the information necessary to contact the parents or guardians of teen Discord users, however our Family Center was designed to provide parents and guardians with tools to better understand how their teens use our service. The Family Center allows parents to get insights into the communities their teen children have joined, including which users their teens have recently added as friends, which users their teen has messaged or called in DMs or GDMs, and which servers their teen has joined or participated in. The Family Center also allows parents to get a weekly email with a high-level summary of the time their teen spends on Discord. The Family Center is designed not just to provide parents and guardians with greater insight into how their teen uses Discord, but to also help parents start conversations with their teen about online safety.

6. **When a person—regardless of age—informs Discord that there is a sexually explicit image of that person being shared on Discord without that person's consent, what is the process that person must go through in order to have that image removed from your platform? What is the average amount of time required for Discord to verify and remove a sexually explicit image after notification?**

   Non-consensual adult intimate media (NCAIM) is not allowed on Discord. A user can report suspected NCAIM to Discord for review. Additionally, we leverage machine learning and artificial intelligence to detect and remove servers devoted to NCAIM. Full-time Discord employees oversee the resolution of cases escalated by our external

partner agents and investigate where NCAIM violations intersect with CSAM and child safety. Our average resolution for NCAIM reports is less than 24 hours.

7. **Discord employs many of the industry standard tools like PhotoDNA and digital "hash" matching. Discord also has engineers dedicated to developing novel methods to detect previously unknown CSAM images. Does Discord attach any identifying marker to nonconsensual images shared on your platform other than those applied by PhotoDNA or similar services? In other words, if a person reports their own nonconsensual image being shared on Discord, what can Discord do to ensure that image cannot be shared in perpetuity?**

   At this time, Discord has not implemented digital hash matching for nonconsensual images shared on the platform in violation of our Community Guidelines. However, we continue to evolve our technological capabilities with regard to enforcement of our Community Guidelines.

8. **The 2022 Thorn Report identified that 17 percent of the minors who use Discord have had a sexual interaction through Discord, and 10 percent believed that the person they were interacting with in a sexual way was an adult. What are you doing to make sure that those interactions are eliminated altogether?**

   Discord has a policy on teen self-endangerment, an issue that we do not take lightly. We want teen Discord users to be able to express themselves freely while also taking steps to ensure that they don't engage in risky behaviors that could endanger their safety and well-being. In order to help teen users stay safe, our policies state that users under the age of 18 are not allowed to send or access any sexually explicit content, and users over 18 are only allowed to post explicit content in spaces that are behind an age-gate or in DMs or GDMs with users who also are 18 and older. We also believe that teens dating online can result in self-endangerment. Under this policy, teen dating servers are prohibited on Discord, and we will take action against users who are engaging in this behavior.

   In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. Among the tools announced as a part of this new initiative are Sensitive Content Filters, which automatically blur potentially sensitive media sent to teens in DMs, GDMs, and in servers. Blurring is enabled by default for teen users in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users (whereas adult users can choose to opt-in to this feature by changing their Privacy & Safety Settings).

   Last fall, Discord also launched Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and

unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, which leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

Senator Ted Cruz
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

I. Directions

Please provide a wholly contained answer to each question. A question's answer should not cross-reference answers provided in other questions.

If a question asks for a yes or no answer, please provide a yes or no answer first and then provide subsequent explanation. If the answer to a yes or no question is sometimes yes and sometimes no, please state such first and then describe the circumstances giving rise to each answer.

If a question asks for a choice between two options, please begin by stating which option applies, or both, or neither, followed by any subsequent explanation.

If you disagree with the premise of a question, please answer the question as-written and then articulate both the premise about which you disagree and the basis for that disagreement.

If you lack a basis for knowing the answer to a question, please first describe what efforts you have taken to ascertain an answer to the question and then provide your tentative answer as a consequence of its reasonable investigation. If even a tentative answer is impossible at this time, please state why such an answer is impossible and what efforts you intend to take to provide an answer in the future. Please further give an estimate as to when Senator Cruz will receive that answer.

To the extent that an answer depends on an ambiguity in the question asked, please state the ambiguity you perceive in the question, and provide multiple answers which articulate each possible reasonable interpretation of the question in light of the ambiguity.

II. Questions

1. **In the last two years, has an employee or commissioner of the Federal Trade Commission (FTC) requested to evaluate or evaluated your data used for training Large Language Models or algorithms or the sources of such data for bias, discrimination, or misinformation?**

   No.

2. **In the last two years, has an employee or commissioner of the FTC sought details regarding your company's measures related to filtering or blocking inputs and**

outputs of a Large Language Model or algorithms?

No.

a. **If yes, has the FTC attempted to coerce or otherwise request you to implement input/output filtering in order to allegedly comply with federal law?**

No.

3. **In the last two years, has an employee or commissioner of the Federal Trade Commission sought to evaluate your company's use of measures, including "prebunking" or "debunking", designed to counteract so called "online misinformation"?**

No.

4. **In June 2022, the FTC released a report titled "Combatting Online Harms Through Innovation." In this report, the FTC discussed how the deployment of AI tools intended to detect or otherwise address harmful online content is accelerating but may never be appropriate as an alternative to human judgment.**

a. **In the context of protecting children from online harms to what extent does your company rely on automated tools to detect online harm vs. human review? Please be specific.**

Discord users are encouraged to report content that violates its policies. Reports concerning the highest harm material, including exploitative content and minor safety, extremism, and cybercrime, are prioritized for review and enforcement. Specialists in Discord's Trust & Safety teams are trained to quickly recognize and handle exactly these types of content. In specific circumstances, reports are escalated to a dedicated investigations team, which is empowered not just to investigate specific infractions, but to conduct wider investigations and refer users to law enforcement agencies as appropriate.

Discord invests in developing innovative technologies, including the use of machine learning and other AI tools, to proactively identify novel child sexual abuse material (CSAM)—images that have not previously been detected and so do not appear in available hash databases. We also invest in new ways to detect the exploitation of minors (commonly referred to as grooming).

i. **Novel CSAM:** Recently, Discord collaborated with industry peers to build and implement a visual safety technology to detect unknown CSAM and AI-generated CSAM with positive results. Effective tools that keep children safe online should be industry standards, not a means for companies to gain a competitive advantage. In recognition of that, as of October, we have made this technology open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online.

ii. **Exploitation of Minors:** Text-based sexual exploitation and extortion of minors is a persistent and growing problem. Last fall, Discord launched Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

   On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, which leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

iii. **Sensitive Content Filters:** This feature automatically blurs potentially sensitive media sent to teens in DMs, GDMs, and in servers. Blurring is enabled by default for teen users in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users.

b. What benefits can AI provide to helping detect and/or stop harmful content to children online?

AI and machine learning can provide significant benefits in helping to detect and combat harmful content online, including content that poses minor safety risks. Discord is committed to leveraging all available tools and technologies as it works to prevent, detect, and respond forcefully to content that harms minors.

AI and machine learning can increase the efficacy of existing tools and technologies and are vital components of the work Discord is doing to halt the spread of CSAM. For example, Discord has invested in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online. By leveraging these tools, Discord's model had an 87 percent increase in recall, i.e. proactively-detected confirmed positive CSAM.

AI- and machine learning-powered technologies can also be leveraged to detect and prevent attempts to exploit minors via text messaging. In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. Among the tools announced as a part of this new initiative are Sensitive Content Filters, which automatically blur potentially sensitive media sent to teens in DMs, GDMs, and in servers. Blurring is enabled by default for teen users in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users. Adult users can choose to opt-in to this feature by changing their Privacy & Safety Settings.

Last fall, Discord also launched Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn, that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think carefully before replying to messages from

strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn, which leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify teenagers about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

Additionally, AI and machine learning play important roles in addressing and attacking these issues at scale. Discord is able to filter, detect, block, and take action on more content today than would ever be possible without these technologies.

c. What does a human reviewer provide that an AI or automated tool cannot? Will we always need some measure of human review in assessing online harms to children?

Human review of potentially violative content is particularly valuable in conducting quality assurance checks. As tools powered by AI and machine learning are deployed on new areas of harm, humans conduct spot checks and audits that guide the development and deployment of the tools. As such, at this time, humans are critical to refining the models, enhancing consistency, and directing consequential decisions around banning accounts and servers.

Humans with subject matter expertise play a critically important role in investigating and actioning particularly nuanced cases involving potential harms to children, such as grooming. Discord devotes significant time to investigating and appropriately reporting such instances to the National Center for Missing and Exploited Children (NCMEC), and human review is a uniquely valuable part of that work.

d. The FTC has sent mixed signals in its enforcement of COPPA. While the Commission emphasizes not over relying on use of automated tools or AI, they have nonetheless found liability for using human review as alternative

signaling overreliance on automated tools. What improvements, if any, should Congress make to clarify the legal tension between use of automated detection tools vs. human review?

As it pertains to potential COPPA enforcement, Congress should ensure that neither automated nor human review can give rise to the "actual knowledge" that underpins a COPPA violation when done in good faith. Discord has policies and procedures in place to investigate and verify a user's age in situations where either the user, within the platform, admits that they are underage or Discord receives a credible report, supported by clear and convincing evidence, that the user is underage.

Outside the context of COPPA, Congress should expressly clarify that companies cannot face liability for: (i) removing content through AI-powered tools and (ii) leveraging human review of CSAM, including for quality assurance and auditing. Enacting carve outs specific to this work will help Discord as well as other companies and vendors build robust detection systems, collaborate, and provide supervision and quality control of those systems.

5. **In 2021, Congress directed the FTC to research and report on how AI can be used positively to detect and combat fraudulent or deceptive content online. Rather than viewing AI as a potential solution to our online woes, the FTC instead issued a report that read more like an indictment of the technology.**

    a. **Please explain whether, in your view, AI can be used to positively detect and combat fraudulent or deceptive content, including the recent use of deepfakes or other scams to harm consumers.**

    Yes, AI can be used as a tool to positively detect and combat such content. Addressing these issues requires a robust system that leverages a combination of machine learning, behavioral heuristics, hashing and matching systems, and other tools—all of which need to work in concert. These models must constantly evolve to address the new challenges presented by the iteration of the technologies used by bad actors and to ensure that the use of AI does not lead to the over-moderation of legitimate content. As such, building tools that effectively combat such content requires collaboration and a significant commitment of resources.

    b. **Has the FTC ever consulted with your company to learn how your company deploys AI to better detect and combat fraudulent or deceptive content? Has**

the DOJ? How about the Federal Elections Commission?

No.

c. **How can Congress empower agencies to use AI positively for the protection of American consumers from fraudulent or deceptive content?**

Discord's use of machine learning and AI to detect and remove content that violates our Community Guidelines suggests that federal agencies could leverage similar technologies to analyze reported fraudulent or deceptive content, provided sufficient safeguards are in place to help ensure accuracy and privacy. To the extent that such uses of AI may require congressional authorization or oversight, this may be an area for congressional action.

6. **Please provide a description of your company's policy regarding the sale or transfer of the data of American users collected on your platform to a third party, including data brokers.**

Discord does not sell user information, nor does it transfer user information to data brokers. Discord uses various service providers, such as cloud storage providers and payment processors, who store and process user information at Discord's direction. Discord may also transfer user information when instructed to do so by the user, such as when a user connects their Discord account to a third-party service.

7. **Has your company ever sold the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the categories of data sold.**

No.

8. **Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the circumstances underlying the basis for such transfer.**

No.

9. **Has your company ever sold the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and the categories of data sold.**

No.

10. **Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and categories of data transferred.**

    Yes. Discord has disclosed limited basic subscriber information such as IP addresses and session logs in response to emergency disclosure requests in emergency situations involving threats to life or serious bodily injury. In the U.S., Discord receives, responds, and communicates with federal and local law enforcement on these matters.

11. **Does your company have a policy to restrict third party use and/or transfer of data collected from users on your platform? Please be specific, including how you enforce such restrictions and whether such restrictions prohibit the sale or transfer of such data to a government agency, including a foreign government agency.**

    Yes. Discord uses contractual language in our data processing addendums, plus employee training, internal access controls, and data governance to restrict the transfer of data.

12. **Between July 4, 2023 and July 14, 2023, was your company contacted by any employee of or contractor for any of the following agencies? Please answer "yes" or "no" for each agency and, if "yes," provide the date(s) of contact and the name(s) of the agency employees or contractors that contacted your company.**

    The answers below reflect requests from the specified federal agencies that were received via Discord's Government Request Portal during the period specified.

    a. **U.S. Department of Health and Human Services (HHS)**

       No.

    b. **National Institute of Allergy and Infectious Diseases (NIAID)**

       No.

    c. **Centers for Disease Control and Prevention (CDC)**

       No.

d. U.S. Food and Drug Administration (FDA)

No.

e. The National Institutes of Health (NIH)

No.

f. U.S. Department of Homeland Security (DHS)

Yes. In 2023, Discord was contacted by the DHS on July 6, 7, 9, 10, 11, 12, and 14.

g. DHS Cybersecurity and Infrastructure Security Agency (CISA)

No.

h. U.S. Census Bureau

No.

i. Federal Bureau of Investigation (FBI)

Yes. In 2023, Discord was contacted by the FBI on July 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, and 14.

j. U.S. Department of Justice (DOJ)

Yes. In 2023, Discord was contacted by the DOJ on July 3, 5, 7, 10, and 13.

k. The White House Executive Office of the President (EOP)

No.

l. U.S. Department of State

No.

13. Is it your company's policy to prevent children under 13 from using your social media app(s) or creating an account?

Yes.

14. In your view, would it be appropriate for school-aged children to spend time on or access your company's social media app(s) during class?

Yes, in some circumstances and for use among study groups and friends. Discord is a real-time communications platform and not a traditional social media application. Many children between the ages of 13-18 used Discord regularly during remote learning, particularly in 2020, to informally collaborate on group projects and work with their classmates on homework assignments.

15. As a parent, would you be concerned if your child were able to access your company's social media app(s) during class via a school network or device?

My children are both under the age of 5 and are not currently allowed to use social media apps under any conditions.

16. In your view, should elementary and secondary schools block students' access to your company's social media app(s) on school networks and devices?

No. Discord is a real-time communications platform and not a traditional social media application. Like other communications tools, Discord can be used effectively for informal collaboration. Discord has been a sponsor of National PTA and collaborates with the organization on digital safety resources and events designed to elevate youth voices and foster shared understanding among teens, caregivers, and educators. Schools are best positioned to determine the access restrictions for their students, and not the providers.

Because users must be at least 13 years old to create an account on Discord, elementary school students are prohibited from using Discord.

17. Do you think that school buses equipped with Wi-Fi should allow children to access your company's social media app(s) via a school bus Wi-Fi network during their rides to and from school?

Discord is a real-time communications platform and not a traditional social media application. Users must be at least 13 years old to create an account on Discord. Like other communications tools, Discord can be used effectively in a variety of settings.

18. As a parent, do you think it is important to supervise your children's internet access?

Yes.

19. As a parent, would you be concerned if your child's school allowed your child to access the internet on an unsupervised basis, such as on your child's bus ride to and from school via the school bus Wi-Fi?

My children are both under the age of 5 and are not currently allowed unsupervised access to the internet.

20. Do you think Congress should require schools, as a condition of receiving broadband subsidies through the Federal Communications Commission's E-Rate program (which funds broadband for elementary and secondary schools), to block students' access to your company's social media app(s) from school-run networks?

No. Discord is a real-time communications platform and not a traditional social media application. Like other communications tools, Discord can be used effectively for informal collaboration in some circumstances.

Because users must be at least 13 years old to create an account on Discord, elementary school students are implicitly prohibited from using Discord.

21. Do you support the bipartisan *Eyes on the Board Act of 2023*, S. 3074?

We agree with the goal of ensuring that children are safe online. However, Discord is concerned that the bill could limit important access to information, collaboration, and learning. We look forward to working with the Committee as they continue to explore these issues.

22. Have you, your company, or any foundation associated with you or your company, donated or contributed funding, equipment, or services to any of the following organizations in the last ten years (CY 2013 to CY 2023)?

a. Education and Libraries Networks Coalition (EdLiNC)

Jason Citron: No.
Discord Inc.: No.

b. Open Technology Institute

Jason Citron: No.
Discord Inc.: No.

c. Consortium for School Networking (COSN)

Jason Citron: No.
Discord Inc.: No.

d. Funds For Learning

Jason Citron: No.
Discord Inc.: No.

e. State Educational Technology Directors Association (SETDA)

Jason Citron: No.
Discord Inc.: No.

f. Schools, Health, and Libraries Broadband Coalition (SHLB)

Jason Citron: No.
Discord Inc.: No.

g. State E-Rate Coordinators' Alliance (SECA)

Jason Citron: No.
Discord Inc.: No.

h. EducationSuperHighway

Jason Citron: No.
Discord Inc.: No.

i. All4Ed

Jason Citron: No.
Discord Inc.: No.

j. Public Knowledge

Jason Citron: No.
Discord Inc.: No.

k. Fight for the Future

Jason Citron: No.
Discord Inc.: No.

l. Free Press

Jason Citron: No.
Discord Inc.: No.

m. Electronic Frontier Foundation

Jason Citron: No.
Discord Inc.: Yes.

n. Benton Foundation or Benton Institute for Broadband & Society

Jason Citron: No.
Discord Inc.: No.

o. Electronic Privacy Information Center

Jason Citron: No.
Discord Inc.: No.

23. For each such donation or contribution described in the prior question, please detail (1) the type of donation or contribution, such as financial donation, goods or equipment, services, etc.; (2) who made the donation or contribution; (3) the recipient organization; (4) the year the donation or contribution was made; and (5) the total value of that donation or contribution.

Discord Inc. donated a total of $252.48 to the Electronic Frontier Foundation (EFF)—these donations were made through our Employee Giving Program, wherein either Discord matches employees' donations to non-profit organizations up to $1,000 per employee per year in time and/or funds, or Discord seeds employees a Deed credit which

they can donate to the non-profit of their choice. Since 2021, Discord has used Deed, a workplace giving and volunteering platform, to administer the Employee Giving Program.

Occasionally, in order to incentivize employees to use the Program, Discord occasionally "seeds" employees a Deed credit that they can then donate to the non-profit of their choice. Discord does not match "seeded" funds.

Through Deed, 4 employees made 19 donations to EFF over three years: $73.60 in 2022, $158.88 in 2023, and $20 in 2024. Because Discord matches employee donations made via Deed up to $1,000 per employee per year, Discord Inc. matched 15 donations to EFF: $25 in 2022, $120 in 2023, and $20 in 2024. Four donations ($87.48 total) represent Deed credit "seeded" by Discord: $48.60 (two donations of $24.30) in 2022, and $38.88 (two donations of $19.44) in 2023. In total, Discord donated $252.48 to EFF.

Senator Josh Hawley
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **Do you allow your children to use social media? If so, please explain under what conditions.**

   My children are both under the age of 5 and are not currently allowed to use social media under any conditions.

2. **Do you believe that children under the age of 18 should be allowed to use social media?**

   Discord is a real-time communications platform and not a traditional social media application. Users must be at least 13 years old to create an account on Discord.

3. **How many individuals does your company employ in Trust & Safety?**

   More than 15 percent of our full-time employees (FTEs) work on safety company-wide at Discord. By way of context, that means our Safety team is larger than our Marketing team. Discord employs 74 FTEs in Trust & Safety and more than 400 external partners and contract agents in Trust & Safety.

4. **How many individuals does your company employ to review content for so-called "misinformation," "disinformation," or "malinformation"?**

   As a real-time communications platform, Discord is not set up for public posts or the viral sharing of information. Therefore, we have prioritized minor safety over the content mentioned. As such, we do not have dedicated employees with their sole focus on misinformation, disinformation, or malinformation.

5. **How many dollars per year does your company spend on salaries for Trust & Safety officers?**

   In 2023, the combined salaries of all full-time employees on the Trust & Safety team at Discord was approximately $11.1 million.

6. **Do you believe that the algorithms your company has developed to sort users' feeds are protected by Section 230 of the Communications Decency Act of 1995? If so, please explain why.**

   Discord is a real-time communications platform that does not algorithmically sort and serve curated feeds.

7. **Do you believe that the algorithms your company has developed to sort users' feeds are expressive speech protected by the First Amendment to the U.S. Constitution? If so, please explain why.**

   Discord is a real-time communications platform that does not algorithmically sort and serve content to user feeds.

8. **Is your company a member of a party, an amicus, or a member of an amicus in *NetChoice, LLC v. Paxton*, No. 22-555 (U.S.), or did your company provide any funds or donations to any party or amicus in that case? If so, please describe the amount of funds or donations made and the context.**

   Discord is an amicus in *NetChoice, LLC v. Paxton*, No. 22-55 (U.S.). Discord's amicus brief can be found here.

9. **Do you believe that the First Amendment to the U.S. Constitution precludes Congress from enacting legislation holding social media companies liable to users for torts they commit?**

   I am not an attorney and I do not believe I have the authority or expertise to answer the question you present here.

10. **Do you believe that companies can be trusted to develop artificial general intelligence (AI) through open-source methods?**

    Discord takes a "safety by design" approach to our work—a process that includes a risk assessment during the product design phase to identify and mitigate potential safety risks. This approach means that our teams pursue responsible product design by studying human behavior and considering the impacts of our product on user safety—before we launch a product into the market.

    Discord has developed AI tools through open-source methods and has made open source new machine learning tools that we helped develop in the child safety space, so that we can contribute to the broader fight against child sexual abuse material (CSAM) online by sharing our successes with industry peers and outside organizations. For example, Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source.

11. **Do you believe the government should play a role in licensing certain artificial intelligence technologies, such as generative AI products?**

    I do not believe I have the authority or perspective to determine the proper role of the

government in licensing artificial intelligence technologies.

12. Do you believe that artificial intelligence represents an existential threat to humanity?

I do not believe I have the authority or perspective to determine whether artificial intelligence represents an existential threat to humanity.

13. Do you think that the development of large language models by Microsoft, Google, Meta, and other large companies raises antitrust concerns?

I do not believe I have the expertise to determine whether development of large language models by large companies raises antitrust concerns.

14. What steps does your company take to make transparent the algorithms by which users are censored, shadow banned, or demonetized?

Discord does not practice censorship or shadow banning. We may demonetize Discord creators who violate our policies, but do so based on human review.

Our quarterly Transparency Reports contain discussions of the kinds of moderation tools we use. In those reports, Discord explains that it relies on various image hashing and machine learning powered tools to detect possible Community Guidelines violations, overwhelmingly in the minor safety context. Discord has also discussed these features in blog posts and other articles on its website, Discord.com. There is no human review for a content moderation decision if a user shares a piece of known CSAM that is identified by PhotoDNA. For other possible violations, the content is reviewed by Discord external partner agents or full-time employees.

There are limited circumstances in which Discord might limit certain functionality for a particular user. Historically, that practice has been limited to accounts that engaged in spam and other inauthentic behavior. For example, if a new account is created on Discord and within minutes it joins dozens of servers and sends hundreds of messages a minute, we may limit the ability for that account to send messages or join new servers. Because this behavior is often the result of individuals or groups who intend to harass and spam users on the service, Discord intentionally limits the extent to which it reveals its methods for detecting such activity.

Discord offers its users only a handful of ways that they may monetize on the service, and those avenues are largely experimental and limited to a small number of total users. Discord does not rely on algorithms to determine whether to limit or remove a user from accessing its monetization features.

15. What steps does your company take to ensure that your company is not disproportionately targeting or censoring conservative voices?

Discord is committed to providing a safe and enjoyable environment for all our users—regardless of their political views—and takes content moderation and safety seriously. To that end, we take pride in the processes, systems, and policies that we have independently developed to guide our approach to content moderation. Discord makes its own decisions on content moderation and safety. To set the baseline for interactions on the platform, Discord publishes Terms of Service and Community Guidelines, outlining what is and is not allowed on Discord. Discord publishes explainers for a wide range of its policies on the Discord Policy Hub, to better make users aware of what the Terms, Guidelines, and policies entail.

We have crafted and honed these public-facing policies with the goal of meeting the needs of our business and users, including their safety and free speech as well as our own free speech rights and business goals. Discord's Community Guidelines are designed to prohibit user behavior that breaks the law or causes serious harm. None of the restrictions in Discord's public-facing policies implicate or regulate political speech unless it otherwise violates the policies. When we do take action to enforce our Terms of Service and Community Guidelines, we strive to be transparent about the nature and quantity of our content moderation efforts. To that end, we publish regular Transparency Reports which quantify the types of actions we have taken and for what category of content.

To identify and action content that violates Discord policies, Discord employs a Trust & Safety team. Discord identifies content and servers that violate its policies in three fundamental ways: (1) through user-generated reports, (2) through proactive moderation by a dedicated team of Discord employees, (3) through Community Moderation, and (4) via a trusted reporter process that allows trusted partners to surface content and reports of violations directly to our Safety team. Proactive monitoring is done by specialist teams trained in evaluating and acting upon certain types of high-harm conduct, including Violent Extremism, Exploitative and Child Sexual Abuse Materials, and Cybercrime. Community moderation allows individual community administrators and moderators to establish rules on top of those Guidelines, to better set the tone for their individual communities, and may remove additional content at their own discretion and based on their own free speech rights. For example, the owner of a server devoted to Green Bay Packers can limit posts by users asserting the superiority of the Dallas Cowboys, and vice-versa. This type of moderation is in addition to, and not instead of, other moderation conducted on the platform. Should a user be removed by a private server moderator for violating server rules (like posting pictures of cats in a server dedicated to pictures of dogs) they can join other cat-related servers or start their own.

Additionally, it is important to emphasize that Discord is a real-time communications platform for friend groups and small communities. It is not a traditional social media platform with posts to the general public that have the potential to go viral or be suppressed based on algorithmic functions or human intervention. Our product is not designed to amplify, promote, or publish public messages to a mass audience in order to increase user engagement and achieve ad sales goals.

16. Do you condemn Hamas' terrorist attacks on the State of Israel on October 7, 2023?

Yes, Discord and I both condemn terrorist attacks on civilians around the globe, including the October 7, 2023 attack by Hamas.

17. What role do you believe social media companies have in promoting or limiting public speech regarding the events of October 7, 2023?

Discord is a real-time communications platform and not a traditional social media site.

18. What investments has your company made in anti-CSAM technology?

Discord uses PhotoDNA to proactively scan images uploaded to our service in order to detect and report CSAM content and perpetrators to the National Center for Missing and Exploited Children (NCMEC), which subsequently works with local law enforcement to take appropriate action. In addition to hash matching, Discord uses internally developed tools (including machine learning techniques) and works with industry partners (including peer companies, non-profits, and researchers) to detect CSAM distribution.

In addition to our use of industry standard tools to detect and remove known CSAM, Discord invests in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online. By leveraging these tools, Discord's model had an 87 percent increase in recall, i.e. proactively-detected confirmed positive CSAM.

We also invest in new ways to detect the exploitation of minors (commonly referred to as grooming). In October 2023, Discord announced Teen Safety Assist, a series of new features designed to keep teens safe on our platform. One of the tools announced as a part of this new initiative is Sensitive Content Filters, which automatically blur potentially sensitive media sent to teens in direct messages (DMs), group direct messages (GDMs), and in servers. Blurring is enabled by default for teen users both in DMs and GDMs with friends and in servers. In DMs and GDMs with non-friends, potentially sensitive media is blocked by default for teen users. Adult users can also choose to opt-in to this feature by changing their Privacy & Safety Settings.

Last fall, Discord also launched Safety Alerts on Senders, a tool designed in collaboration with the child safety non-profit Thorn that alerts teen users on Discord when a conversation is potentially unwanted. This new feature is on by default for teen users and helps determine when a DM sent to a teen from an unfamiliar user should trigger the delivery of a safety alert to the teen recipient. The alert encourages teens to think

carefully before replying to messages from strangers, allows teen users to block new and unfamiliar senders, and provides a link to safety tips for how teens can best protect themselves online.

On February 6, 2024, Discord announced Safety Alerts in Chat, a new feature also developed in collaboration with Thorn that leverages machine learning algorithms to analyze messages between adults and teenagers. This product aims to identify messages that may be indicative of grooming behaviors. Upon detection, it will notify a teen user about these potentially harmful messages and offer a suite of safety tools, including options to block and report the sender, along with providing advice on dealing with an unwanted situation and encouraging the teen user to take a break from the conversation, reach out to Crisis Text Line for live help, or visit NoFiltr.org for more resources and safety tips. This new feature will start rolling out in the coming weeks.

As part of our extensive fight against CSAM, we continued to deepen our partnership with the Tech Coalition. One of the most exciting parts of our partnership with the Tech Coalition this year was the launch of Lantern, a first-of-its-kind signal sharing program for companies to enhance and strengthen how we detect attempts to sexually exploit and abuse children and teens online. Discord joined other industry partners like Google, Meta, Roblox, Snap, and others to share signals about activity we detect that violates our policies and utilize signals shared by other companies to surface violating activity on their platforms.

Our largest acquisition to date has been on a machine learning safety company called Sentropy. Discord is committed to making heavy investments in safety and it's reflective in how we prioritize safety.

19. Have you read the Fifth Circuit's opinion in *Missouri v. Biden*, No. 23-30445?

No.

20. Do you dispute any factual findings in the Fifth Circuit's opinions or the district court's opinions?

I have not read *Missouri v. Biden* and am not in a position to dispute its factual findings. However, I've been informed that Discord is incorrectly characterized as a "social media company" in the district court's opinion and that there are no references to Discord in the Fifth Circuit opinion.

21. Does your platform continue to receive requests from federal agencies to censor or promote certain content?

From time to time, Discord receives notices from federal agencies regarding illegal content that violates our community guidelines and we action accordingly.

22. What steps do your platforms take to verify and enforce age restrictions?

When an individual attempts to register a new account, Discord requires the individual to enter their date of birth through a neutral age gate. If the individual provides a date of birth indicating that they are under 13 years old, Discord does not permit the individual to register an account and instead presents a notice screen informing the individual that they are "unable to register." If the individual selects "Back to Login," they are shown the typical registration landing screen and if they select "Register," they are once again shown the "unable to register" notice without being provided an opportunity to make up a new birthdate or attempt to re-register. Discord provides additional information to the public about why the service asks for individuals' date of birth in a published Trust & Safety article.

Discord also takes steps to confirm that registered users are 18 years of age or older when they attempt to join an "age-restricted" server or channel. As outlined in the Community Guidelines, users must be age 18 or older to participate in communities containing adult content on Discord and servers and channels organized around adult content or age-restricted themes (e.g., violent content) must be labeled as age-restricted. When a user first attempts to join an age-restricted server or channel, Discord implements an extra layer of protection, through an age-restricted warning or gate.

Discord also investigates and verifies a user's age in situations where either (i) the user, within the platform, admits that they are underage or (ii) Discord receives a credible report, supported by clear and convincing evidence, that the user is underage. For instance, Discord sometimes receives communications allegedly from a parent or guardian asserting that their underage child has a Discord account. Discord's Trust & Safety team takes these reports seriously and follows processes for investigating such reports and, where appropriate, verifying the user's age and disabling the user's account if we determine that the user is under 13.

Finally, Discord undertakes efforts to identify a user's age when a user whose account has been banned for being underage submits an appeal. Discord employs a rigorous verification process for investigating and arbitrating such appeals with an emphasis on giving the user (or their parent) an opportunity to definitively prove that the user is age 13 or older, while leaving properly executed bans in place.

23. In cases where a child's safety is at risk, how does your company collaborate with law enforcement? What information or assistance is provided?

Discord provides user information to law enforcement when we receive valid legal process or when there are emergency situations when we possess a good faith belief that there is an imminent risk of serious physical injury or death. We also may proactively report to law enforcement when we inadvertently come across information that appears to pertain to the commission of a crime, and are especially likely to proactively disclose this information in situations which appear to be high harm. When we come across

information that relates to child abuse material, we report that to NCMEC.

24. Do you believe there is any expressive value in CGI or AI generated CSAM?

Discord does not allow on our platform any content or activity that sexualizes a minor, including AI-generated CSAM. We do not find value in such images, and we prohibit such images entirely. When such imagery is found, it is removed and reported to NCMEC, which subsequently works with local law enforcement to take appropriate action. This policy is designed to ensure that the content does not proliferate and that the sexualization of children in any context is not normalized by bad actors.

25. Do you believe that CGI or AI generated CSAM is protected by the First Amendment to the U.S. Constitution?

I am not an attorney and I do not believe I have the expertise to answer this question. As such, I have no position on the constitutional status of CGI or AI generated depictions of CSAM. I understand that Discord has the right to prohibit such content on our platform. We do not allow on Discord any content or activity that sexualizes a minor, including AI-generated CSAM. When such imagery is found, it is removed and reported to NCMEC, which subsequently works with local law enforcement to take appropriate action. This policy is designed to ensure that the content does not proliferate and that the sexualization of children in any context is not normalized by bad actors.

26. What measures does your platform take to ensure that children only see age-appropriate advertisements?

Discord historically has not sold advertising and instead builds premium features that users can choose to purchase. For example, users can buy subscriptions to products like Nitro, which gives users the ability to upload larger files, stream video in high definition, and personalize the user experience by using custom emoji and digital stickers in conversations with others. Discord users can also purchase Server Boosts, which allow users to share these perks with members of a server to which they belong. These features are designed to improve the experience of using Discord, not to maximize engagement by keeping users on the app.

27. Will you commit to setting up a compensation fund for those who have been harmed by your platform?

We take our responsibility to our users seriously. Discord's commitment to safety is reflected in our investments in improving online safety and the way we build our products. It starts with the technology we build and the expertise we rely on to manage it. We also work closely with law enforcement when appropriate and support legislative efforts to ensure they can hold perpetrators of crimes accountable. For example, we support elements of the SHIELD Act that create new criminal offenses related to the non-consensual distribution of intimate visual depictions.

28. Does Discord have access to and provide copies of users' direct message communications to federal law enforcement upon request?

Discord is required to produce content data upon receipt of legal process reviewed by a judge or magistrate. Discord is generally required to produce content data upon receipt of a judicially-reviewed legal process, such a search warrant, subject to Discord's legal objections.

29. Has Discord ever received a request from an executive agency to monitor or promote certain kinds of content?

We are not aware of any communications to Discord from any executive agency seeking the monitoring or promotion of certain kinds of content. We have, on occasion, been invited by executive agencies to participate in summits, roundtables, and listening sessions on policy topics.

30. Some generative AI models, most notably Midjourney, use Discord as their user interface. Does Discord monitor the images created in this way?

Images generated through AI models are monitored via the same visual detection system as all other messages and images on the platform. As a result, we scan the attachments stored on Discord for CSAM via hash matching. For instances where the user submits images to modify with AI, the input images are scanned for CSAM as well.

Additionally, images generated through AI models are monitored via the new visual detection tools used to detect unknown and AI-generated CSAM, which Discord built in collaboration with industry peers.

31. Has Discord created policies to monitor the use of these models within the platform?

Discord's platform policies and community guidelines apply to all content on Discord, including prompts made to generative AI models and the content produced by these models.

32. Does Discord have Trust & safety employees dedicated to monitoring the content created by generative AI that is hosted on Discord?

Discord's Minor Safety and Exploitative Content (MSEC) team works to detect and action this type of content. Additionally, our machine learning teams are focused on developing new ways to tackle novel and AI-generated CSAM.

Senator Thom Tillis
Senate Judiciary Committee
Questions for the Record
Jason Citron, Co-Founder and CEO, Discord Inc.

1. **Twenty-one is the minimum age to purchase highly regulated adult products such as alcohol, tobacco, and nicotine. Nevertheless, there is a proliferation of user-generated content posted on social media sites featuring underage use of these products.**

   Recently, some have proposed banning these age-restricted products due in part to the user-generated content being available on your respective platforms. Surely, banning these products cannot be the answer. However, we must do more – your company must do more – to shield underage audiences from exposure to this content.

   Therefore, as the content moderator of these platforms, what policies do you have in place, and what more can you do, to prevent this type of user-generated content from reaching underage audiences? How do you respond to requests to pull this content from your sites when deemed inappropriate for underage audiences?

   Discord's Community Guidelines provide that users are not permitted to organize, promote, or engage in any illegal behavior, such as the buying, selling, or trading of dangerous and regulated goods. Dangerous and regulated goods include but are not limited to: firearms; explosives; tactical gear; imitation firearms; illicit drugs; marijuana; alcohol; and tobacco.

   However, while users are prohibited from illegal behavior related to dangerous or regulated goods (i.e., minor users attempting to buy alcohol on Discord), Discord does allow general discussions and content about such goods (i.e., a server in which adult users discuss their favorite breweries) provided that the content is posted behind an age-gated channel or server as outlined in our Age-Restricted Content Policy. Discord's Community Guidelines also provide that users must be age 18 or older to participate in communities containing adult content on Discord and servers and channels organized around adult content or age-restricted themes must be labeled as age-restricted. When a user first attempts to join an age-restricted server or channel, Discord implements an extra layer of protection, through an age-restricted warning or gate.

2. **Public reports conclude that drug cartels use social media like TikTok, META, X, Snapchat, and others to plan, organize, and communicate in real-time. These communications coincide directly with criminal activity.**

What are your companies doing to crack down on cartel coordination? Specifically, in the recruitment of children to commit crimes or assist in the sale/distribution of illicit drugs?

Discord takes immediate action against any identified criminal activity, employing advanced detection methods, including network analysis of bad actors and keyword matching, to proactively identify and address spaces suspected of engaging in the illegal sale of substances. We continuously monitor our platform for illicit activities and take immediate action when we find it.

3. **What steps does your platform take to proactively remove, delist, and ban any posts, users, websites, and advertisements associated with the sale and distribution of fentanyl and other illicit drugs?**

Discord enforces its Dangerous and Regulated Goods Policy prohibiting the sale and distribution of dangerous or regulated goods, including illicit drugs. We employ a dedicated Cybercrime team equipped with sophisticated tools and methodologies to proactively detect and dismantle networks involved in such illegal activities. This includes the use of keyword scanning to identify and take action against entities selling or distributing drugs and other illicit substances. Approximately 80 percent of our interventions are initiated proactively, with the remaining 20 percent resulting from user reports. Furthermore, we actively engage in partnerships with industry peers and governmental bodies, including the DEA, to continuously refine our detection capabilities and response strategies, ensuring we are at the forefront of combating the digital facilitation of drug trafficking.

4. **One area of growing concern is the sale and distribution of fake or counterfeit vaping devices online, particularly in connection with so-called Delta-8 THC. Counterfeit vapes, many coming from China, have unsafe and even potentially deadly chemicals. They have caused hospitalizations and death. What are your platforms doing to combat this problem?**

Discord strictly prohibits the sale of fake or counterfeit vaping devices, recognizing the severe health risks they pose, including those associated with Delta-8 THC products. We actively enforce Discord's Dangerous and Regulated Goods Policy through proactive detection methods, including advanced algorithms and keyword scanning, to identify and remove violative servers and users. In addition to our technology-driven approaches, we also encourage user reports of suspected counterfeit items. Upon identification, we take down these spaces and take appropriate actions against the users involved.

5. **What are the main impediments your platform encounters in identifying all fentanyl and illicit drug advertisements posted to your platform(s) automatically? Please describe any circumstances in which you do not or cannot apply detection technologies against content transmitted on your platform(s).**

   Discord employs various proactive methods to detect illicit drug advertisements, including those related to fentanyl. The development and implementation of advanced machine learning (ML) models to enhance detection accuracy require considerable staffing to manage and refine these models, which poses challenges to mid-sized companies. Additionally, the high rate of false positives often associated with such detection models can present a significant challenge to automated enforcement.

6. **How many posts, users, websites, and advertisements have you removed, delisted, and banned per year for the sale and distribution of fentanyl and other illicit drugs? How many per year? Have you seen an increase in illicit drugs being advertised to children on your platform(s)?**

   In our ongoing efforts to combat the sale and distribution of illicit drugs, including fentanyl, on our platform, we have taken the following enforcement actions regarding illegal goods over the past three years:

   - In 2021, we removed 52,995 users and 777 servers.
   - In 2022, we removed 20,244 users and 456 servers.
   - In 2023, we took action against 4,578 users and 463 servers.

   Discord cannot disaggregate this data further to just illicit drugs, but these actions encompass illicit drugs, weapons, and other illegal goods. Discord's active enforcement of our Dangerous and Regulated Goods Policy, particularly our proactive detection methods that leverage advanced algorithms and keyword scanning, has enabled the platform to better identify and remove violative servers. As these enforcement statistics suggest, we have observed a downward trend in the number of drug-related servers on our platform.

7. **Are there any other roadblocks or impediments that you face in addressing fentanyl and illicit drug advertisements on your platform(s), and working with law enforcement on such matters? If yes, what are they? If no, how many cases have been transmitted to law enforcement and DEA?**

   Enhanced cooperation with law enforcement agencies, including the DEA, and increased

industry collaboration have aided in improving our capabilities through shared intelligence. Discord is developing its protocols for more effective information sharing which will result in greater capability to transmit drug-related information to law enforcement agencies.

8. **How do you work with organizations, advocates, and experts focused on drug prevention and addiction recovery to adapt your products and operations to keep up with the illicit drug crisis — including working with parents that have lost children due to lethal drugs bought online?**

   In addition to our work with the National PTA (discussed in greater detail below) to create digital safety resources and events for teens, caregivers, and educators, Discord is also a member of the Family Online Safety Institute (FOSI) and supports ConnectSafely, a non-profit organization that educates people about online safety, privacy, security, and digital wellness.

9. **What are the total number of meetings that your company has had with parents to address online safety concerns? Can you provide the total number of meetings over the last three years? Please separate this last question's answer by number per year.**

   Discord's work with parents and guardians, as well as parent-led organizations, is vital to our child safety efforts. For example, Discord has met regularly with National PTA nearly every-other week since late 2021. Since 2022, Discord has been a sponsor of National PTA. Discord collaborates with National PTA on digital safety resources and events designed to elevate youth voices and foster shared understanding among teens, caregivers and educators. The PTA Connected: Build Up and Belong program, sponsored by Discord, facilitates technology discussions between teens and their parents and caregivers about online safety topics. In 2023, 25 local PTAs were selected to receive $1,250 each, sponsored by Discord, to host a PTA Connected: Build Up and Belong program. Discord also sponsored grants for this program in 2022.

10. **In 2022, then National Center for Missing & Exploited Children (NCMEC) received over 32 million reports of Child Sexual Abuse Material (CSAM). Reports of online sex crimes to the CyberTipline are growing exponentially year by year. Out of those 32 million reports, how many did your platform submit to NCMEC?**

    In 2022, Discord submitted 58,179 reports to NCMEC. In 2023, we submitted 165,478 reports to NCMEC. These metrics can be found in our Transparency Reports.

11. **There is concern that this number is going to fall dramatically this year because of**

the adoption of end-to-end encryption, not because the problem is going away. How will your company track and address this issue moving forward?

Text messages on Discord are not end-to-end encrypted and we commit to keeping teen users' text messages unencrypted. The features Discord deploys to help keep teens safe on our service, such as our Teen Safety Assist feature, are designed to identify conversations with teen users that could present a risk to their safety, and the effectiveness of our teen safety features would be undermined by encrypting the text messages of teen users.

Additionally, Discord is committed to working with NCMEC to detect and remove CSAM on our platform and the broader internet through initiatives like the Technology Coalition's Project Lantern.

12. **Has your platform seen an increase of suspected online child sexual exploitation-CSAM over the past few years? If so, what do you believe is the driving factor on why it's happening on your platform?**

Discord continues to invest in developing innovative technologies to proactively identify CSAM, including AI-generated, photorealistic CSAM and novel CSAM—images that have not previously been detected and so do not appear in available hash databases. While we are unable to discern a trend in the prevalence of suspected online child sexual exploitation-CSAM, Discord does publish quarterly Transparency Reports that detail our enforcement actions with regard to violations of our Community Guidelines, including violations of our child safety policy, and provide greater insight into the work we do to keep users safe.

13. **What are some new tools or strategies that your platform has implemented to identify CSAM? How closely does your platform work with NCMEC?**

Discord uses PhotoDNA to proactively scan images uploaded to our service in order to detect and report CSAM content and perpetrators to NCMEC, which subsequently works with local law enforcement to take appropriate action. In addition to hash matching, Discord uses internally developed tools (including machine learning techniques) and works with industry partners (including peer companies, non-profits, and researchers) to detect CSAM distribution.

In addition to our use of industry standard tools to detect and remove known CSAM, Discord invests in developing innovative technologies to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available

hash databases. Discord collaborated with industry peers to further develop CLIP, a machine learning image detection and identification algorithm, into a specialized version designed to detect unknown and AI-generated CSAM. In recognition of the industry-wide need to develop and deploy tools to effectively identify novel CSAM, as of October we have made Discord's CSAM-trained CLIP implementation open source, so that we can share our successes with other organizations—without cost—and contribute to the broader fight against CSAM online. By leveraging these tools, Discord's model had an 87 percent increase in recall, i.e. proactively-detected confirmed positive CSAM.

As part of our extensive fight against CSAM, we continued to deepen our partnership with the Tech Coalition. One of the most exciting parts of our partnership with the Tech Coalition this year was the launch of Lantern, a first-of-its-kind signal sharing program for companies to enhance and strengthen how we detect attempts to sexually exploit and abuse children and teens online. Discord joined other industry partners like Google, Meta, Roblox, Snap, and others to share signals about activity we detect that violates our policies and utilize signals shared by other companies to surface violating activity on their platforms.

Discord also works closely with NCMEC, utilizing their hash database in our internal models as well as developing technology to detect novel CSAM and opening that technology up to other companies through Project Lantern.

14. What resources or help does your platform provide to victims of CSAM? Does your platform work with local victim groups and professionals?

Discord's approach to child safety is driven by expert research that draws from diverse perspectives, including survivors. For example, we collaborated with the researchers from Thorn to ensure we were taking a research backed approach to designing Teen Safety Assign feature—a tool to help Discord identify messages that may be indicative of grooming behaviors—to ensure that it reflected the lived experience of teens.

Additionally, through our participation in Tech Coalition and other convenings led by groups like WeProtect Global Alliance, we are committed to constantly learning and evolving our approach to better support young people and those who may be victims of online sexual abuse. For example, earlier this year three Discord employees attended the Multi-Stakeholder Forum to Combat the Financial Sextortion of Young People, where discussions about developing safety tools and approaches centered the experiences of young people and victims.

Discord has also participated in the 2022 WeProtect Global Alliance convening which featured a presentation from the Brave Movement, a survivor-centered global movement

fighting to end childhood sexual violence.

15. What are the top technical hurdles your company faces in combatting CSAM?

Discord is proud of its work to develop and make open source new tools and technologies to better detect, remove, and halt the spread of not just known CSAM, but AI-generated CSAM and other text-based threats to child safety online, such as grooming. As discussed above, we have invested in a tool to proactively identify novel CSAM—images that have not previously been detected and so do not appear in available hash databases. We have also invested in new ways to detect the exploitation of minors (commonly referred to as grooming).

Effective tools that keep children safe online should be industry standards, not a means for companies to gain a competitive advantage. We will continue our work to build new tools and share our successes with other companies and organizations.

16. There seem to be competing views on how to regulate algorithms. Some suggest that more transparency is needed, while others want more privacy. Can you provide your perspective on whether more or less transparency is needed when it comes to algorithms?

As an initial matter, it is important to clarify that Discord is a real-time communications platform. On Discord, content from other users is displayed in real time in the order in which it is posted. Content on private servers can only be viewed by users who have joined the server by invitation or request. Discord users can be invited to specific servers by other users, or they may find servers that they want to join via a keyword search available through Discord's "Explore Discoverable Servers" function.

Discord has experimented with features that leverage algorithms; for instance, highlighting a limited amount of user-generated content to community members. To be clear, none of these features promote or encourage virality of posts (like public trending topics or stories), because the content being surfaced is user-specific and limited to users who are already members of the community where the feature is active. The technology powering the features is content agnostic. That is, user engagement is the critical element in determining what is highlighted; the underlying content is not a factor.

Discord is not a service designed to maximize engagement by an algorithm picking and choosing the content users see. Rather, Discord emphasizes real time interaction and connection among friends. Moreover, where Discord offers features that leverage algorithms to suggest content, the company prioritizes giving users choice, for instance

the ability to opt-in or opt-out of such features.

Discord provides transparency around its limited use of algorithms while still preserving user privacy. Beyond that, I cannot speak to algorithms that Discord does not deploy and are otherwise outside my scope of experience and expertise.

17. Do you believe that large companies and platforms like yours can use algorithms to stifle innovation or small businesses?

Discord is an innovator, not a large company. We have fewer than 1,000 employees.

18. What do you believe is the role of government in regulating algorithms? What, if any, unintended consequences would there be if Congress gets involved?

I do not believe I have the experience or perspective to determine the proper role of government in regulating algorithms.

19. Are you aware of your platform using surveillance advertisements to target children (anyone under the age of 18) with specific ads? If so, in your opinion, how can this be mitigated?

No, Discord does not engage in this behavior..

20. Beyond surveillance advertisements, are there any other algorithmic-based practices being implemented that are particularly detrimental to children? In your opinion, how can this be mitigated?

Discord is a real-time communications platform that does not rely on algorithmic virality to drive usership. Discord does not have awards, quantification, or down-vote or up-vote features that incentivize users to post certain types of content over others. Discord does not use algorithms to recommend or amplify content to users from servers of which they are not already a member. As such, Discord does not have the same expertise in this area as some of its peer companies in order to opine on potential mitigation.

21. Are you aware of any surveillance advertisements or algorithms that are used to target children, specifically to promote drugs and the sale of narcotics?

No.

Senate Judiciary Committee
Questions for the Record for Mr. Evan Spiegel
CEO and Co-Founder of Snap, Inc
Submitted February 28, 2024

Sen. Dick Durbin (D-IL)

1. For each year from 2019 to 2023, please provide the following:

a. the total number of users on your platform;

Approximate Global Average Annual Daily Active Users (DAU) is shown below.

	Approximate Global Avg. Annual DAU
2019	205M
2020	245M
2021	300M
2022	354M
2023	400M

b. the total number of users under the age of 18 on your platform;

	Approximate Global Avg. Annual DAU Aged 13-17
2019	34M
2020	41M
2021	50M
2022	58M
2023	60M

c. the estimated number of users under the age of 13 on your platform;

Snapchat is not intended for use by people under the age of 13. When we learn an account user may be under 13 years old, we investigate and remove the account if warranted. We therefore do not track the number of Snapchatters who may be under age 13.

d. the number of users of your platform under the age of 18 who opted-in to your Family Center tool and were linked with a parent or guardian's account?

Approximately 200,000 parents use Family Center and about 400,000 teens have linked their account to their parents using Family Center.

e. your company's annual revenue;

2019 – approximately $1,716 million
2020 – approximately $2,507 million
2021 – approximately $4,117 million
2022 – approximately $4,602 million
2023 – approximately $4,606 million

f. your company's annual budget for trust and safety;

2019 - approximately $39 million
2020 - approximately $54 million
2021 - approximately $131 million
2022 - approximately $164 million
2023 - approximately $135 million

g. your company's annual budget to address online child sexual exploitation;

See above totals in (f), as this area of focus is incorporated in our trust and safety budget.

h. the total number of employees working to address trust and safety;

Year	Total No. Performing Safety and Moderation Work (As of 12/31/23)
2019	763: 79 (FTE), 684 (CW)
2020	1,218: 99 (FTE), 1,119 (CW)
2021	3,051: 148 (FTE), 2,903 (CW)
2022	2,593: 124 (FTE), 2,469 (CW)

2023	2,226: 163 (FTE), 2,063 (CW)

i. the total number of employees working to address online child sexual exploitation.

Same answer as (h) above.

2. How did your company determine that 13 was the appropriate age for a child to begin using your platform?

In compliance with the Children's Online Privacy Protection Act (COPPA), Snapchat is not a service directed to children under the age of 13. When registering a Snapchat account, users are required to accept Snap's Terms of Service, acknowledge the Privacy Policy, and provide their date of birth. Users need to be at least 13 to create a Snapchat account, and the registration process is blocked if a user is under the age of 13. If we become aware that a Snapchat user is under the age of 13, we will terminate that user's account and delete the user's data. We also implement a safeguard that prevents 13-17 year old Snapchat users with existing accounts from updating their birthday to an age of 18 or above.

3. What legal obligation does your company have in the United States to ensure that your platforms are safe for children before they are launched?

As described in response to question 2, COPPA provides a legal framework for children's access to internet services. Snapchat is not a service directed to children under age 13. Users need to be at least 13 to create a Snapchat account, and the registration process is blocked if a user enters an age under 13. If we become aware that a Snapchat user is under the age of 13, we will terminate that user's account and delete the user's data. In 2014, Snap entered into an agreement with the FTC to establish and maintain a robust privacy program, which, among other things, involves pre-launch reviews of Snap's features and functionality. The mechanisms in place as part of these processes play a critical role in promoting safety when designing Snapchat features for all users, including our teenagers. Over the years, we have expanded the FTC privacy-by-design framework to also incorporate safety-by-design, and have created a holistic framework for privacy-and-safety by design, that considers a wide range of laws and regulations.

4. For users under the age of 18,

a. what are the default privacy settings for their accounts?

Snapchat was designed with privacy and safety in mind for all users, and we apply a high privacy bar for all users, including teens. For example, friend lists are private by default, and we don't offer settings to change that. Contact settings are limited to friends and phone contacts only – and can't be expanded. Location sharing is off by default, users under 18 are unable to create a Public Profile on Snapchat, and we don't allow under 18s to change their age to an age over 18.

b. what limitations are placed by default on content these users can access, content that will be directed toward them, and individuals they can communicate with?

We want Snapchat to be safe for everyone, and we offer extra protections for minors to help prevent unwanted contact and provide an age-appropriate experience. We have default content settings for teens that limit their exposure to shocking, suggestive, or other sensitive content on our broadcast surfaces. We also use age-gating to ensure that any advertising content is age-appropriate.

As a foundational safeguard, we've designed our service to require direct communication between friends to be opt-in, meaning people have to proactively choose who they communicate with. Friend lists are private on Snapchat, which not only reduces social pressure but also limits the ability of predators to find a person's friends on Snapchat. Snapchat's default "Contact Me" settings are set to friends and phone contacts only for all accounts, and can't be expanded. If a minor receives a friend request from someone they don't share a mutual friend with, we provide a warning before they start communicating to make sure it is someone they know. As a result, approximately 90% of friend requests received by minors on Snapchat are from someone with at least one mutual friend in common. Our goal is to make it as difficult as possible for people to be contacted by someone they don't already know.

c. can they change their default settings without the awareness of their parent or guardian, or without the consent of their parent or guardian?

Yes, users under 18 can change their default settings without the consent of their parent or guardian. However as part of our ongoing enhancement and evolution of our Family Center tool for teens and parents, we plan to introduce a new feature to give parents visibility into their teen's privacy settings.

d. in 2023, how many changed their default settings?

We do not possess this data.

5. If the default settings are different for users aged 16 and 17 than they are for users under the age of 16, please explain why your company takes this position, how this position was developed, and whether any company personnel voiced objections to or raised concerns about this position.

Default settings are the same for all users under 18.

6. What studies, research, summaries, or data does your company have reflecting the efficacy of its parental controls and child safety measures? Please provide these studies, research, summaries, or data.

In developing Snap's parental resource tool, Family Center, we took a comprehensive approach to find the right balance of features that gave parents insight to who their teens were interacting with on Snapchat, while still respecting the privacy and autonomy of teens. Our research highlighted how important it was to provide useful tools for parents that improve visibility into their teen's activity on our service and expand education on safety best practices, to safeguard teen privacy and limit overly invasive tools, and to be mindful of different family dynamics across cultures and marginalized groups. This research included:

- Direct user feedback, submitted by our users though our support site
- Focus groups that interviewed parents and teens separately and together to get their feedback on which tools would be most useful
- User research studies that surveyed parents and teens on their perception of safety on Snapchat, and gauged their awareness of Family Center and parental tools on other services
- Benchmark studies that survey parents' sentiment towards Snapchat and parental tools and how this changes over time
- Feedback sessions with dozens of online safety experts, including academics, researchers, safety experts, members of parent groups, and NGO leaders to inform our design of Family Center features
- Deep dive discussions with our Safety Advisory Board on their recommendations for our parental tools

7. Concerning international law,

> **a**. what steps have your company and its subsidiaries taken to comply with the European Union's Digital Services Act?

At Snap, privacy, safety, and transparency have always been core to how we operate. We have protections in place for all members of our community and offer additional safeguards for our teenage Snapchatters. Our long-standing values are aligned with the principles of the European Union's Digital Services Act (DSA) and we share their goals to create a safe online environment. We made a number of changes to our service, including (1) giving Snapchatters the ability to control the content they're shown, (2) a new notification and appeals process for content or account removals, (3) updates to our advertising practices, and (4) appointing compliance officers.

> **b**. what steps has your company and its subsidiaries taken to comply with the United Kingdom's Online Safety Act?

The United Kingdom's Online Safety Act (OSA) is not yet in effect. However, our long-standing privacy and safety practices, and work relating to other compliance efforts, provides a strong foundation for new requirements that will come into effect under the OSA in the coming years. We are incorporating OSA requirements in our risk assessments and privacy-and-safety by design review process.

> **c**. what steps has your company and its subsidiaries taken to comply with Australia's Online Safety Act?

Snap's longstanding robust privacy and safety practices provided a strong foundation for compliance with Australia's Online Safety Act, and we did not need to make significant changes to our best practices to comply. The requirements of Australia's Online Safety Act are incorporated as part of our ongoing privacy-and-safety by design review process.

> **d**. if those laws create a safer, healthier online experience for kids on your platforms, do you commit to implement these changes in the United States? If not, why not?

Snap has in place protections for all members of our community and offers additional safeguards for our teenage Snapchatters, including our users in the United States, and all users benefit from safety enhancements we've made as part of compliance with the laws mentioned above. Our long standing values are aligned with the principles of the European Union's Digital Services Act (DSA) and other safety and transparency laws, and we share their goals to create a safe online environment. We also support the Kids Online Safety Act, introduced in the Senate. We have already implemented many of its provisions such as mandating appropriate default privacy settings, providing safeguards for teens (including additional privacy settings that protect their privacy and offer them control over their experience), offering parental tools, and limiting the collection and storage of personal information.

8. In 2022, Snapchat was used by 90% of U.S. residents aged 13-24, making Snapchat a prime platform for predatory users to target children. According to Snap's Terms of Service convicted sex offenders are not permitted to use its services. Despite this policy, registered sex offenders have been found to regularly make accounts and further victimize and exploit children using Snapchat.

For example, in August 2023, a registered sex offender with prior convictions in Illinois and Michigan was sentenced to 25 years in prison for luring a 15-year-old girl over Snapchat and engaging in sexual activity with her. Similarly, in July 2023, a registered sex offender was arrested in Texas and charged with child trafficking to engage in sexual conduct. The offender used Snapchat to meet his potential victims online. What steps does Snap take to enforce its policy against registered sex offenders and ensure sex offenders are actually kept off of your platform?

These crimes are vile and abhorrent. We do not tolerate such accounts on Snapchat and we have in place policies allowing us to take swift action and help bring predators to justice. If we learn an account is alleged to be used by a registered sex offender, our teams immediately investigate and take action as appropriate. We receive tips from law enforcement, news media, and other partners, and we act on these referrals. Upon confirming that an account is used by a registered sex offender, the account is immediately disabled and the device is blocked.

9. Snap uses photo recognition technologies to detect CSAM on its platform, but this technology can only detect known CSAM. As a result, newly-created CSAM often goes undetected on Snapchat. As a platform where users are overwhelmingly sharing new pictures and videos, how is Snap working to prevent the use of its platform for the creation and trafficking of CSAM?

We ban sharing nude images of anyone under 18 and want to protect our community from the devastating consequences that can come if this content falls into the wrong hands. If we find this content, we report it to NCMEC. In 2023, for example, more than 225,000 of the approximately 690,000 NCMEC reports Snap submitted originated from users reporting content to us. We empower users to report this content with easy-to-use in-app reporting tools, and we are exploring additional technical solutions for detecting this content.

Sen. Lindsey Graham (R-SC)

1. Do you support S. 1207, the bipartisan EARN IT Act? Why or why not?

We support the Senate Judiciary Committee's efforts to better protect young people online, and are supportive of the Earn It Act's underlying goals. As currently drafted, we have some concerns with this bill. The broad liability provisions in the Earn It Act could be used to sweep up services, like ours, in a wave of civil litigation that distracts from our core safety mission, despite our adherence to best practices. A FOSTA-SESTA approach would allow for liability against services that are truly bad actors—those who are affirmatively trying to help criminal bad actors on their service, or willfully turning a blind eye to known instances of illegality—while still protecting services that are working tirelessly to do their best to stop bad actors.

2. What measures are you taking to prevent and address sextortion, including financial sextortion, on your companies' platforms?

This conduct is illegal and abhorrent, and we take active steps to identify and prevent such exploitation as well as empowering users to identify and report suspicious contact.

We combat this criminal activity by:

- Making it difficult for strangers to search for, find, or contact teens.
- Warning teens if they receive a friend request from another person who isn't a mutual friend or phone contact.
- Keeping friend lists private and preventing teens from having public profiles — which helps prevent criminals from using Snapchat to target a teen's friends
- We proactively detect accounts attempting to engage in sextortion and disable them. When we remove sextortion-related content, we also retain it for an extended period of time in case law enforcement wants to follow up with a valid legal request.
- We've added an in-app reporting option specifically for sextortion ("They leaked/are threatening to leak my nudes.") to make it easier for our community to report such abuse to us. When we receive such a report, we take action quickly – usually within 30 minutes.
- We also educate teens about the dangers of this type of crime and urge them to use our in-app reporting tools.

This is very important because young people may feel afraid to report the problem to their parents or caregivers.

a. What methods are in place to detect and disrupt this type of abuse in real time?

Please see response to question 2.

3. Please provide the committee statistics on how long it takes your company to respond to various types of legal process from law enforcement?

Currently, we respond to most legal process within two to three weeks. We respond to most requests for voluntary disclosure of data in instances involving imminent threat to life or serious physical injury within thirty minutes.

4. Do you notify your users when law enforcement serves subpoenas/summons for subscriber information and specifically requests not to notify the subscriber/user?

We do not notify users when law enforcement provides a nondisclosure order or when we believe an exceptional circumstance exists, such as cases involving child exploitation, or the threat of imminent death or serious bodily injury, such as the sale of fentanyl. When law enforcement does not provide a nondisclosure order, we will follow up to see if they plan to obtain a nondisclosure order before we take steps to notify the user.

> **a**. If you notify the subscriber, how long do you wait until notification goes out?

Snap will not notify a user until we have confirmation from law enforcement that they understand the user will be notified, that law enforcement does not intend to seek a non-disclosure order and we have given law enforcement the opportunity to let us know whether the case does not fall within one of the exempted categories listed in the answer above.

> **b**. Are you aware that by notifying the subscriber about a law enforcement subpoena for their subscriber information that you are jeopardizing critical evidence that could be erased before law enforcement can serve warrants?

The existing legal framework (the Stored Communications Act) permits law enforcement to submit a preservation request, and Snap does not notify users upon receipt of a preservation request. Snap complies with such requests to preserve account data for up to 180 days as authorized by law. A preservation is a snapshot in time of a user's available data, including available basic subscriber information, metadata, geolocation data, and content (*e.g.*, Chats, Stories, and Memories) and is held in an offline file where it cannot be accessed by the user.

We do not provide user notice when providing notice is prohibited by a court order or by other legal authority; or when we believe an exceptional circumstance exists, such as cases involving child exploitation, or the threat of imminent death or serious bodily injury, such as the sale of fentanyl. When law enforcement believes that critical evidence would be erased as a result of user notification, they commonly seek and are granted a non-disclosure order. Due to the use of non-disclosure orders and Snap's policy to exempt cases involving the exceptional circumstances noted above, fewer than 5% of account holders for whose accounts data was sought were notified of legal process between June and December 2023.

c. Would your company agree to a 90-day non-disclosure to subscribers to allow law enforcement ample time to secure proper legal process?

A nondisclosure period is already provided in existing law—the Stored Communications Act—upon submission of a valid court order by federal or state law enforcement. We do not provide user notice when providing notice is prohibited by a court order or by other legal authority. We also choose not to provide notice when law enforcement provides us with information indicating an exceptional circumstance exists, such as cases involving child exploitation, or the threat of imminent death or serious bodily injury, such as the sale of fentanyl.

5. Do you actively seek out and incorporate feedback and insight from survivors of online sexual exploitation to improve your trust and safety policies and practices and to prevent and disrupt child sexual abuse material (CSAM) production and distribution on your platform? Can you provide examples?

a. If not, please explain.

One of our Safety Principles at Snap is that we strive to be victim- and survivor-informed in all our efforts, including child sexual exploitation and abuse. We have the honor of representing the technology industry on the international Policy Board of the WeProtect Global Alliance, which includes representation of a disclosed survivor. These board members' contributions have impacted our thinking on a broad range of issues. We have also engaged with survivors directly and indirectly at CSEA-related conferences and events via our Safety Advisory Board, and we regularly engage with experts and nonprofits in this area.

6. During our hearing, you testified that you collaborate with parents and parent organizations to create mechanisms to keep children safe online. Please elaborate and cite examples of your company's work with non-employee parents and parent organizations.

We consulted with parents as part of a series of focus groups in developing our parental resource tool, Family Center, and with nearly 40 child safety experts to seek feedback and input before the launch of Family Center. We are also active Family Online Safety Institute members, support ConnectSafely.org, and collaborate with Protect Young Eyes. In addition, we work closely on national public awareness campaigns with Song For Charlie, a group started by Ed and Mary Ternan, who lost their son to fentanyl poisoning, and regularly consult members of our Safety Advisory Board, who engage with parents and caregivers globally.

7. Why does your company have the age limit of 13 years old for a user to sign up for an account?

In compliance with the Children's Online Privacy Protection Act (COPPA), Snapchat is not a service directed to children under age 13. When signing up for Snapchat, users are required to accept Snap's Terms of Service, acknowledge the Privacy Policy, and provide their date of birth.

Users need to be at least 13 to create a Snapchat account, and the registration process is blocked if a user is under the age of 13. If we become aware that a Snapchat user is under the age of 13, we will terminate that user's account and delete the user's data. We also implement a safeguard that prevents younger Snapchat users with existing accounts from updating their birthday to an age of 18 or above. Specifically, if a Snapchat user between the age of 13 to 17 attempts to update their year of birth to reach an age over 18, we will prevent the change.

a. Why not younger or older?

Please see response to question 7.

8. How many minors use your platform? How much money does your company make annually from these minors?

In 2023, there were approximately 60 million 13-17 year old daily active users on average globally. In 2023, Snap's global revenue from minors was approximately $437M.

9. What percentage of your employees work on trust and safety and how much money does your company invest annually in trust and safety?

Approximately 20% of our team, including both employees and contingent workers, works on safety – including our moderation teams, trust and safety teams, law enforcement teams and more. In 2023, we spent approximately $135 million on personnel costs for this team.

10. It is sometimes challenging for law enforcement conducting criminal investigations to determine the true identity of a person behind a name on social media or other online platforms, and whether an online identity is an actual person. What are you doing to validate the true identity of users – or the fact that a user is a human – when they create an account on your platforms?

We believe Snapchat can help people build stronger and healthier relationships with the most important people in their lives. To support that goal, we work everyday to help our community have a safe experience on our service – it's not only the right thing to do, it's essential to our mission. A key part of this work is collaborating with law enforcement, and we have a global Law Enforcement Operations team that works around the clock to support law enforcement investigations.

While messages on Snapchat delete by default – designed to reflect the nature of real-life conversations between friends – we can preserve available account information and content upon receipt of a valid law enforcement request.

To help build confidence that a new user is in fact a legitimate person, we prompt users during the account registration flow to verify their phone number or email address provided. We are also vigilant in regard to accounts being created through suspected automation (e.g. spam actors). We conduct risk analysis during account creation (and also post account creation) using continuously curated technical and behavioral signals to tell apart inorganic or undesired traffic coming on to our service.

11. Is your company using safety technology to detect and prevent live video child sexual abuse on your platforms and apps that allow users to stream or share live video? If not, please explain.

Snapchat does not have video capabilities that would permit users to live-stream or broadcast live video.

> **a**. Has your company tested that or similar technology? If not, are you developing similar technology to address child sexual abuse in live video?

See response to Question 11.

12. How are you measuring if your trust and safety policies, practices, and tools are effective in protecting children from sexual abuse and exploitation on your platform?

> **a**. What specific metrics or key performance indicators do you use?

Our semi-annual transparency reports are an important tool to hold ourselves accountable and share information and updates on our efforts to combat violating content and accounts on our service. You can find our latest Transparency Report here. Our main Transparency Report includes statistics relating to the Violative View Rate and our California transparency page links to our new California Terms of Service Report. This report, in turn, includes a range of safety-related metrics, including providing the Violative View Rate for individual categories of harms, as well as the Violative Viewer Rate which discloses how many viewers saw the content as a percentage of all active users over a period of time, both for the U.S. and globally.

13. Is your company using language analysis tools to detect grooming activities? If not, please explain.

No, we do not use language analysis tools to detect grooming. We believe there are other tools like network analysis that can detect and prevent grooming activities without compromising the privacy of user communications.

We are not aware of grooming-detection technology based on language analysis that has achieved the sufficiently high levels of precision required in light of the privacy considerations,

but we are actively exploring other metadata and content signals that can help uncover bad actor accounts.

a. What investments will your company make to develop new or improve existing tools?

We have already deployed features that limit the discoverability of minors via Search and Quick Add to people they have multiple friends and contacts in common with, which is a deterrent to grooming because it makes it difficult for strangers to request to connect with minors. We will continue to invest in more advanced strategies like this to make it even more difficult for strangers to find minors on the service. Additionally, we released an "in chat warning" to minors, surfacing block and report options at the top level, if they become friends with someone outside of their existing friend network. We will expand on this feature to incorporate more risk signals and provide even stronger warnings to minors in certain cases (e.g. if the other user has been reported before).

14. What resources have you developed for victims and survivors of abuse on your platforms?

As part of our overall effort to prioritize the mental health of Snapchatters, in 2020, we launched Here For You, a proactive in-app support system for Snapchatters who may be experiencing a mental health or emotional crisis by surfacing resources from expert organizations when Snapchatters search on a range of mental health-related topics, including eating disorders, anxiety, stress, depression, hate speech, bullying, suicidal thoughts, and more. When our Trust & Safety team recognizes a Snapchatter in distress, they will forward self-harm prevention and support resources, and notify emergency response personnel as appropriate. The resources that we share are available on our global list of safety resources, and these are publicly available to all Snapchatters.

15. What is your response to requests for content removal from CSAM survivors and other members of the public?

We take prompt action when we receive reports of CSAM, generally responding within 15 minutes. We also prioritize working with experts on programs like NCMEC's Take It Down and the Stop NCII database which gives us even greater access to hashes of illegal content that we can then leverage to proactively identify and remove offending content from Snapchat.

16. Some call Snapchat "dangerous by design," given the platform's disappearing photo feature, which makes it easy for predators to hide their crimes and giving naïve users a false sense of security. In which app functions do you scan for child sexual abuse material (CSAM)?

The sexual exploitation of any young person is horrific, illegal, and against our policies.

We prevent the distribution of CSAM in three key ways:

- We use cutting edge technology – PhotoDNA and CSAI Match – to proactively identify known CSAM photos and videos uploaded to Snapchat and report them to NCMEC.
- When we enforce against illegal and abusive content, we retain it for an extended period so that it is available if law enforcement want to follow up with a valid legal request.
- We take urgent action when we receive user reports of CSAM, generally responding within 15 minutes.

17. What safety messaging does Snap provide to its younger users around online safety, especially as it relates to online enticement and financial sextortion?

In 2023, we launched four new "Safety Snapshot" episodes in-app to help raise awareness and educate our community about various online sexual-related risks, namely: sexting and nudes, financial sextortion, child sex trafficking, and child online grooming for sexual purposes. Safety Snapshot is our official in-app channel for online safety and privacy advice and guidance. These episodes seek to speak to teens in a relatable language and style, and offer links to helpful resources. All four of the episodes were reviewed by experts at the National Center for Missing and Exploited Children (NCMEC) prior to release.

We have also provided users on Snapchat with safety messaging regarding the risks of fentanyl poisoning and counterfeit prescription drugs. We developed and made available an in-app education portal called "Heads Up" that distributes content from expert organizations such as Song for Charlie, Shatterproof, the CDC and the Substances and Mental Health Services Administration. If someone on Snapchat searches for drug-related keywords, Heads Up will show relevant educational content designed to dissuade engagement and ultimately prevent harm to our community.

Additionally, in partnership with Song for Charlie, we developed a video advertising campaign that has been viewed over 260 million times on Snapchat, and rolled out a new national filter that raises awareness of the dangers of fentanyl and counterfeit pills and directs Snapchatters to the Heads Up educational portal.

Finally, Good Luck America, a Snap Original news show, produced a special edition series of episodes devoted to educating our community about the fentanyl crisis.

18. How does Snap age assure and verify its users to ensure young children are not accessing its Platform?

In our Terms of Service, Privacy Policy and other documentation, we make clear that Snapchat is intended for users 13 years old or older. When registering a Snapchat account, users are required to accept Snap's Terms of Service, acknowledge the Privacy Policy, and provide their date of birth. Accordingly, the registration process is blocked if a user declares that they are under the age of 13. If we become aware that a Snapchat user is under the age of 13, we will terminate that

user's account and delete the user's data. We also implement a safeguard that prevents younger Snapchat users with existing accounts from updating their birthday to an age of 18 or above. Specifically, if a Snapchat user between the age of 13 to 17 attempts to update their year of birth to reach an age over 18, we will prevent the change.

Sen. Chuck Grassley (R-IW)

Directions

Please answer each question to the fullest possible extent. If your platform is unable to answer a particular question or does not have requested data, explain why. Each question refers to your company in addition to any corporate affiliates, including parent and subsidiary companies.

1. Current law requires that a provider of a report of suspected CSAM to the National Center for Missing and Exploited Children's (NCMEC) CyberTipline preserve "any visual depictions, data, or other digital files that are reasonably accessible and may provide context or additional information about the reported material or person" for a minimum of 90 days. 18 U.S.C. 2258A(h)(1-2). The recent explosion of suspected abuse has presented unprecedented challenges for law enforcement to follow up on leads before companies discard or delete essential data and information. There is nothing preventing tech companies from preserving relevant material beyond the statutorily-mandated 90-day period.

> **a.** How long does Snap voluntarily preserve and retain data contained in and related to its reports to the CyberTipline?

Snap voluntarily preserves and retains all available data on accounts reported to the CyberTipline for approximately 180 days, twice the legally mandated period. In addition, when we enforce against illegal and abusive content, we retain it for an extended period so that it is available if law enforcement want to follow up with a valid legal request.

> **b.** The massive influx of reports to the CyberTipline naturally results in law enforcement entities having to conduct and finish investigations beyond 90 days of an initial report to the CyberTipline. Retaining relevant information for longer periods could significantly advance law enforcement's ability to thoroughly investigate leads. If Snap only preserves and retains this information for the minimum 90-day period, why does it do so when preserving this data longer could significantly enhance and prolong law enforcement's ability to investigate and prosecute child predators?

Snap already retains all available data on accounts reported to the CyberTipline for approximately 180 days.

> **c.** Please confirm if Snap stores and retains the following information relating to reports to the CyberTipline:
>
> > **i.** IP addresses

Yes, assuming this question is asking whether this category of data is preserved in connection with accounts reported to the CyberTipline.

ii. Screen Names

Yes, assuming this question is asking whether this category of data is preserved in connection with accounts reported to the CyberTipline.

iii. User Profiles

Yes, assuming this question is asking whether this category of data is preserved in connection with accounts reported to the CyberTipline.

iv. Associated Screen Names (by IP address and associated emails)

It is unclear what information the question seeks but Snap preserves a record of usernames and display names for accounts reported to the CyberTipline.

v. Email addresses

Yes, assuming this question is asking whether this category of data is preserved in connection with accounts reported to the CyberTipline.

vi. Geolocation data

Yes, assuming this question is asking whether this category of data is preserved in connection with accounts reported to the CyberTipline.

d. If Snap does not retain or store any of the above types of information in question (c), please explain why.

N/A

e. Please list any other information Snap retains and preserves for law enforcement purposes not listed above in question (c).

Snap preserves and retains available account information on accounts reported to the CyberTipline for 180 days. In addition to the data points listed above, this includes additional subscriber information, available communications content, and stored images and videos, among other data points.

f. Does Snap flag screennames and associated email addresses to suspected accounts that violate Snap's terms of service?

When Snap files a NCMEC report, the report includes the account's username and display name, as well as subscriber information, including email address, if applicable.

2. How does Snap prioritize urgent requests for information from law enforcement and what is Snap's response time to urgent requests?

Snap maintains a 24/7 emergency disclosure request form available to law enforcement globally for cases involving an imminent threat to life or serious physical injury. Snap responds to most such requests in under thirty minutes.

3. What is Snap's average response time to service of legal process from law enforcement for CSAM-related information?

Snap responds to most legal process, including legal process pertaining to CSAM-related information, within two to three weeks.

4. In 2023, the tech industry as a whole slashed more than 260,000 jobs. And in the first four weeks of this year, another 25,000 jobs were cut.

a. For each year, between 2018 and 2023, how many U.S. based employees did you have at Snap?

Year	Total No. of US Employees
2018	FTE: 2,298
2019	FTE: 2,378
2020	FTE: 2,786
2021	FTE: 3,965
2022	FTE: 3,686
2023	FTE: 3,731

i. Of these employees, how many were sponsored on H-1B visas?

Year	Total No. of US Employees on H-1B Visas
2018	323

2019	345
2020	410
2021	564
2022	446
2023	468

ii. For each year, between 2018 and 2023, how many H1-B visa applications did Snap submit?

Year	Total No. of H-1B Visa Petitions Submitted (including amendments and extensions)
2018	173
2019	254
2020	226
2021	359
2022	314
2023	248

b. For each year, between 2018 and 2023, how many employees based outside the U.S. did you have at Snap?

Year	Total No. of Non-US Employees Globally (incl. China)	Total No. of Employees in China
2018	FTE: 586	FTE: 49
2019	FTE: 817	FTE: 45
2020	FTE: 1,077	FTE: 47
2021	FTE: 1,696	FTE: 68
2022	FTE: 1,602	FTE: 60
2023	FTE: 1,558	FTE: 68

i. Of these employees, how many were based in China?

Please see response to question 4.b.

c. For each year, between 2018 and 2023, how many employees in total did Snap terminate, fire, or lay off?

Year	Total Employee Involuntary Attrition Globally	Total Employee Involuntary Attrition - US
2018	FTE: 324	FTE: 297
2019	FTE: 72	FTE: 67
2020	FTE: 37	FTE: 26
2021	FTE: 53	FTE: 40
2022	FTE: 1,266	FTE: 868
2023	FTE: 337	FTE: 162

i. Of these employees, how many were based in the United States?

Please see response to question 4.c.

ii. Did Snap fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?

We do not possess this information, however, we have provided the total number of employees sponsored on H1-B visas by the end of each year in response to question 4(a)(i), and total petitions filed in response to question 4(a)(ii).

iii. Were any duties and/or functions previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?

Please see response to question 4.c.ii.

d. For each year, between 2018 and 2023, how many employees performing work related to child safety did Snap terminate, fire, or lay off?

Year	Total Involuntary Attrition of Safety & Moderation Employees
2018	FTE: 14
2019	FTE: 0
2020	FTE: 1
2021	FTE: 1
2022	FTE: 26
2023	FTE: 4

i. Of these employees, how many were based in the United States?

Year	Total No. of U.S. Involuntary Attrition of Safety & Moderation Employees
2018	FTE: 11
2019	FTE: 0
2020	FTE: 1
2021	FTE: 1
2022	FTE: 17
2023	FTE: 2

ii. Did Snap fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?

We do not possess this information, however, we have provided the total number of employees sponsored on H1-B visas by the end of each year in response to question 4(a)(i), and total petitions filed in response to question 4(a)(ii).

iii. Were any duties and/or functions (specifically relating to child safety) previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?

Please see response to question 4.d.ii.

iv. How have layoffs impacted Snap's ability to protect children on its platforms?

Snap's reduction in force did not impact our ability to do safety-related work, nor our commitment to safety-related initiatives.

v. Does Snap have any plans to increase staff responsible for child safety operations or otherwise optimize its child safety operations?

We are constantly assessing our staffing needs to ensure that we are appropriately staffing these core and important initiatives.

5. On January 30, 2024, the Tech Transparency Project (TTP) published an article on their website called, "Meta Approves Harmful Teen Ads with Images from its Own AI Tool". In summary, TTP, using Meta's "Imagine with Meta AI" tool generated inappropriate images such as young people at a pill party or other vaping. These images with text were submitted to Facebook as advertisements targeting users between ages 13-17 in the United States. TTP reported that Facebook approved the advertisement, despite it violating its own policies, in less than five minutes to run on the following platforms:

Facebook, Instagram, Messenger, and Meta Quest. Meta. Over the course of a week, TTP submitted the advertisements with the same end result: Facebook approving them. TTP reported that they canceled these advertisements before their scheduled publication, but it illustrated the repeated failures of Facebook to properly moderate content. This is just one example of what other non-government organizations and others have uncovered across social media platforms.

a. How often a month do Snap employees conduct quality checks on Snap's policies and safeguards for child accounts?

Snap employees conduct quality checks regularly, including as part of daily audits of our enforcements (which also includes reviews of our policy) and weekly policy calls with our Policy team and 3rd party vendors who support Snap's content moderation efforts. A cross-functional group also meets on a regular basis to discuss safety efforts on Snap and child safeguards. We also conduct adversarial testing work to mimic ways bad actors may attempt to abuse our product and take those learnings to continue to improve our safeguards.

b. In which departments, components, or units of the company does Snap have staff dedicated to performing this type of work?

We believe that safety is a shared responsibility. We have teams in Product, Engineering, Operations, Policy, and Legal that support safety.

c. How many employees make up these departments, components, or units?

We have over 2,200 employees and contingent workers that support safety efforts at Snap. The majority of these personnel support trust and safety and content moderation.

d. If a violation is found, what action is taken, and how quickly is action taken?

Our dedicated Trust and Safety team works 24/7 to review reports and take appropriate action—in the vast majority of cases, we respond to reports and concerns within hours of receiving a report. Please refer to our Transparency Report that provides a breakdown of turn around times for actioning harmful content.

6. Social media companies claim they are investing in company components dedicated to safety, and that their platforms are safe for children. However, children continue to be exploited daily across these platforms.

a. What have Snap's revenue and profit figures been for the last three years (2021-2023)? Please provide figures broken out per year. Do not provide percentages.

Revenue:
2021 – $4,117 million
2022 – $4,602 million
2023 – $4,606 million

Profit/Loss:
2021 - $488 million net loss
2022 - $1,430 million net loss
2023 - $1,322 million net loss

b. How much has Snap spent in advertising for the last three years (2021-2023), broken out per year?

FY21: approximately $221 million
FY22: approximately $309 million
FY23: approximately $269 million

c. How much of Snap's resources spent on advertising has been devoted to advertising Snap's safety initiatives and efforts for the last three years (2021- 2023), broken out per year?

Much of our efforts to educate users about safety initiatives are through in-app notifications and education. When we offer new safety features, we utilize notifications to help Snapchatters understand the feature. We also have an in-house produced Snapchat channel dedicated to safety.

Safety Snapshot covers a range of safety topics such as sextortion, bullying, drugs, and other harmful behavior and educates Snapchatters on what they can do to protect themselves.

In 2022, with the launch of Family Center, Snapchat's parental settings tool, we dedicated $100,000 for media spend to fund advertising awareness for this initiative. This was largely to educate parents (who may be less familiar with our service) about a new tool to help them understand how their teens are using Snapchat.

d. To get an understanding of how your company has invested and plans to invest in its components dedicated to child safety functions, what are the annual budgets for Snap's child safety-related components for the last three years (2021-2023)?

2021 - approximately $131 million
2022 - approximately $164 million
2023 - approximately $135 million

e. What is the current anticipated (2024) budget for Snap's child safety-related components?

2024 Expected - approximately $142 million

f. Provide the number of staff employed in Snap's child safety-related components for the last three years (2021-2023).

Year	Total No. Performing Safety & Moderation Work (As of 12/31/23)
2021	3,051: 148 (FTE), 2,903 (CW)
2022	2,593: 124 (FTE), 2,469 (CW)
2023	2,226: 163 (FTE), 2,063 (CW)

g. How much is that compared to Snap's other components for the same period? (Please provide a breakout per year. Do not provide percentages.)

Year	Total No. of Snap FTEs and CWs (As of 12/31/23)	Total No. Performing Safety & Moderation Work (As of 12/31/23)

2021	11,129: 5,661 (FTE), 5,468 (CW)	3,051: 148 (FTE), 2,903 (CW)
2022	10,618: 5,288 (FTE), 5,330 (CW)	2,593: 124 (FTE), 2,469 (CW)
2023	10,620: 5,289 (FTE), 5,331 (CW)	2,226: 163 (FTE), 2,063 (CW)

h. How many staff are currently employed in Snap's child safety-related components?

We believe that safety is a shared responsibility. We have teams in Product, Engineering, Operations, Policy, and Legal that support safety. We have over 2,200 employees and contingent workers that support safety efforts at Snap. The majority of these personnel support trust and safety and content moderation.

i. What are the roles, responsibilities, and functions of Snap's child safety-related components?

Snap has many core teams across the company working on Child Safety, including those that work across safety of all our features. These teams include Product, Engineering, Operations, Policy, and Legal.

j. Are any other components responsible for the monitoring of CSAM on Snap's platform(s)?

As mentioned, Snap has many core teams across the company working on Child Safety, including those that work across safety of all our features. When our Trust & Safety team identifies CSAM content, they quickly remove the content and enforce against the appropriate account(s).

k. What, if any, third parties does Snap employ or contract with to address CSAM material on its platforms?

In addition to review by Snap employees, over the last three years, Snap has contracted several third parties to assist in review of CSAM material. These included Accenture, Telus, and Oddacious. Additionally, Snap contracts with an intelligence vendor that helps identify violative content on other services and across the internet (and dark web) implicating Snapchat accounts. We then promptly review those reports, investigate the accounts, and remove offending content.

i. What are the roles and responsibilities of these third parties?

Their responsibilities include an initial review of user safety reports as defined by our internal moderation processes and then, if needed, escalation to our full-time employees as instructed in our processes.

ii. What is the breakdown of cost per third party over the last three years (2021-2023)?

These costs are included in the overall Trust & Safety and Content Moderation Budget. Those include:

2021 - approximately $131 million (includes Vendor cost of ~$98 million)
2022 - approximately $164 million (includes Vendor cost of ~$112 million)
2023 - approximately $135 million (includes Vendor cost of ~$86 million)

7. Of all reports sent by Snap to the National Center for Missing and Exploited Children, how many reports were self-generated from victim users for the last three years (2021- 2023)? Please provide the actual number of self-generated reports in addition to the total number of reports (including those that were not self-generated). In addition, please provide a break-down of the self-reporters by age.

Snap is providing the total number of NCMEC reports submitted for 2021-2023 as well as a breakdown of reports that were initiated by user reporting. We do not have a breakdown of user reports by age.

Year	Global NCMEC Cybertips	Number of Global NCMEC Cybertips that were initiated by user reports
2023	690,000	234,600
2022	551,000	38,570
2021	512,000	97,280

8. What is Snap's policy or protocol with respect to law enforcement accessing user data and subsequent notification to users of law enforcement accessing their data?

We do not notify users when law enforcement provides a nondisclosure order or when we believe an exceptional circumstance exists, such as cases involving child exploitation, or the threat of imminent death or serious bodily injury, such as the sale of fentanyl. When law enforcement does not provide a nondisclosure order, we will follow up to see if they plan to obtain a nondisclosure order before we take steps to notify the implicated user.

a. Do certain crimes such as drug trafficking or child exploitation affect Snap's decision to notify a user whose data is accessed by law enforcement?

Yes, we waive our user notice policy when we believe an exceptional circumstance exists, such as cases involving child exploitation, or the threat of imminent death or serious bodily injury, such as the sale of fentanyl.

b. Do certain requests such as a subpoena or search warrant affect Snap's notification protocol? If so, what are they?

Snap's user notice policy does not differ from what is stated above for subpoenas or search warrants.

c. If Snap does notify users of law enforcement accessing their data, why does Snap find this necessary?

We do not provide user notice when providing notice is prohibited by a court order or by other legal authority; or when we believe an exceptional circumstance exists, such as cases involving child exploitation, or the threat of imminent death or serious bodily injury, such as the sale of fentanyl. When law enforcement believes that critical evidence would be erased as a result of user notification, they commonly seek and are granted a non-disclosure order. Due to the use of non-disclosure orders and Snap's policy to exempt cases involving the exceptional circumstances noted above, fewer than 5% of account holders for whose accounts data was sought were notified of legal process between June and December 2023.

In these circumstances where law enforcement does not believe that there is a risk that evidence will be destroyed and thus has not sought a non-disclosure order and where no exceptional circumstance exists based on the vulnerability of the victim or threat of imminent harm, we allow affected users a period of time to challenge the legal process in court.

9. The National Center for Missing and Exploited Children has indicated that reports from social media companies tend to lack actionable information causing law enforcement to be burdened with incomplete information. How comprehensive are Snap's reports to NCMEC? What challenges is Snap experiencing on the collection of user data and other information to include in its reports to NCMEC? What actions is Snap taking to make its reports more comprehensive and useful to law enforcement?

The safety of our Snapchatters, and specifically child safety, is a top priority for Snap. Child sexual exploitation is abhorrent and illegal, and we act quickly when we identify such conduct. Snap's CyberTips submitted through NCMEC's API are comprehensive and include data in the fields provided by NCMEC through its submission API. Snap is regularly in touch with staff from NCMEC and requests feedback regarding CyberTips submitted to NCMEC. When NCMEC raises questions about particular CyberTips or Snap's processes, teams within Snap quickly review and, as appropriate, make changes to our NCMEC submissions. In the past year,

Snap has made numerous changes to our NCMEC submissions processes based on this regular feedback from NCMEC, as well as from members of law enforcement charged with investigating these horrific crimes.

10. The CyberTipline estimates that nearly half of the reports from Snap are lacking essential information, and most critically, child victim identifier information.

Snap's systems are designed to include child victim identifier information (such as the victim username, email address, and phone number) whenever appropriate under Snap's policies. Certain CyberTips do not contain this information by design because there is at times no party to list in the "victim" category. For instance, the most common scenario is when a report implicates a reporting adult user and a reported minor user who is purportedly advertising their own sexual content for sale (in that case, the minor is the reported party, so we do not separately populate their information in the victim fields). The same is true if two adults share CSAM that is proactively detected – in that instance Snap does not know the identity of the child victim.

We endeavor to make all Snap CyberTips actionable by including available relevant information and we actively seek out feedback from NCMEC and members of law enforcement in furtherance of that goal. We also include victim account information when available and in scenarios in which the victim is an account holder.

> **a.** Of the reports submitted to NCMEC by Snap, how many of them lack information identifying child victims?

Not every CyberTip includes victim information. For example, when two adults share known CSAM, the victim identity may be unknown and thus would not contain information identifying the child victim. In all instances in which we become aware that a report incorrectly omits victim information, and we are in possession of such information, we work to retract and resubmit those CyberTips with the information included.

> **b.** Why can't Snap provide child victim identifiers to help law enforcement rescue children who have been exploited on its platform?

Snap does provide child victim identifiers, when appropriate, in CyberTip reports along with additional information, such as the email or phone number associated with the account. We continue to seek input from NCMEC and law enforcement about the actionability of our CyberTips and how to improve them.

11. Saved CSAM images and screenshots, derived from conversations between child predators and exploited children operating their child accounts allow child predators to spread CSAM. For conversations involving child accounts, is it possible to disable a recipient's ability to save or screenshot CSAM messages?

When Snap identifies CSEAI on our service, we remove the content and enforce against the relevant account, and we report the account to NCMEC. While it is not possible to disable a recipient's ability to save or screenshot particular messages, Snapchat notifies a user if another party to a conversation has saved or screenshotted any part of their conversation. The ability to screenshot is a functionality that exists at the operating system level rather than the app-level.

Sen. Josh Hawley (R-MO)

1. Do you allow your children to use social media? If so, please explain under what conditions.

Only one of our children is 13 years old, the minimum age for many online services. We do not allow him to use social media. We allow him to use Snapchat.

2. Do you believe that children under the age of 18 should be allowed to use social media?

Yes, although we believe their experience should be different from an adult experience, which is why we offer a different content experience for teens than adults and also give parents and guardians visibility into their teens' Snapchat experience through Family Center.

3. How many individuals does your company employ in Trust & Safety?

We have over 2,200 employees and contingent workers that support safety efforts at Snap. The majority of these personnel support trust and safety and content moderation.

4. How many individuals does your company employ to review content for so-called "misinformation," "disinformation," or "malinformation"?

Snap does not employ individuals for the dedicated purpose of reviewing content relating to any particular type of harm, such as various forms of information manipulation. Any potential violations of our Community Guidelines are reviewed by our content moderation teams and/or trust and safety teams.

5. How many dollars per year does your company spend on salaries for Trust & Safety officers?

At least 20%, including both employees and contingent workers, of our company works on safety – including our moderation teams, trust and safety teams, law enforcement operations teams and more. For 2023, total spend on salaries for all team members working on trust and safety, product safety, safety engineering, and content moderation and review is approximately $135 million.

6. Do you believe that the algorithms your company has developed to sort users' feeds are protected by Section 230 of the Communications Decency Act of 1995? If so, please explain why.

Yes, we believe Section 230 covers the algorithms used to rank and display content to users. This interpretation is supported by judicial rulings from several federal courts of appeal around the country.

7. Do you believe that the algorithms your company has developed to sort users' feeds are expressive speech protected by the First Amendment to the U.S. Constitution? If so, please explain why.

Yes, courts have consistently held that First Amendment protections apply to the way online services disseminate or curate third-party content.

8. Is your company a member of a party, an amicus, or a member of an amicus in NetChoice, LLC v. Paxton, No. 22-555 (U.S.), or did your company provide any funds or donations to any party or amicus in that case? If so, please describe the amount of funds or donations made and the context.

Snap is a member of NetChoice. Snap did not contribute to any party or amicus in support of the NetChoice LLC v. Paxton case.

9. Do you believe that the First Amendment to the U.S. Constitution precludes Congress from enacting legislation holding social media companies liable to users for torts they commit?

Yes. Snap supports thoughtful regulation designed to minimize harms that may result from the use of online services. Snap does not believe that, consistent with the First Amendment, liability can attach to companies whose only role in a tort that one user commits on another is the transmission of information and facilitation of communication. Attaching liability to online services that are conduits of such speech would force online services to drastically cut back on the ability of users to communicate online, with a significant chilling impact on constitutionally protected speech.

10. Do you believe that companies can be trusted to develop artificial general intelligence (AI) through open-source methods?

Snap does not open-source our artificial intelligence technology. We believe it is important to monitor AI systems and their outputs to ensure they are being used and evolved safely.

11. Do you believe the government should play a role in licensing certain artificial intelligence technologies, such as generative AI products?

Snap believes that a risk-based approach to technology policy is appropriate. There may be narrow, high-risk categories of AI technology for which a licensing regime could prove beneficial; however, detailed information would be needed to assess the virtue of such an approach.

12. Do you believe that artificial intelligence represents an existential threat to humanity?

Artificial intelligence represents a field of technology that is yielding powerful, transformational advancements in human potential. Like other paradigm-shifting scientific achievements, its potential for harm merits serious attention, just as its potential benefits should inspire steadfast cultivation.

13. Do you think that the development of large language models by Microsoft, Google, Meta, and other large companies raises antitrust concerns?

Snap believes competition is critical for safeguarding American technological leadership in the world. Large language models are expensive to develop which may make it difficult for smaller companies to access and benefit from their potential.

14. What steps does your company take to make transparent the algorithms by which users are censored, shadow banned, or demonetized?

We take several steps to ensure that our content moderation policies and practices are transparent to our users. First, we publish our acceptable use policies for public review, which include our Community Guidelines, our Advertising Policies and Commercial Content Policies, and our Content Guidelines for Recommendation Eligibility. We also provide Snapchatters with an overview of how we moderate content, enforce our rules, and enable users to appeal our decisions. Through these resources, we aim to provide transparent guidance about what we do and do not permit to afford users a clear understanding of how we distribute user generated content on our platform.

Second, we provide notice to our users, and opportunities to appeal moderation decisions. When content is removed we provide notice to our users that it has been deleted for violating our Community Guidelines. We also inform users when content is restricted from broad distribution in Spotlight or on their Public Profiles, and tell them why. When an account is deleted for violating our Community Guidelines, the user is notified and afforded an opportunity to appeal.

15. What steps does your company take to ensure that your company is not disproportionately targeting or censoring conservative voices?

We apply our Community Guidelines consistently, regardless of the account owner. Those guidelines expressly note that our rules "apply to all content (which includes all forms of communication, like text, images, generative AI, links or attachments, emojis, Lenses and other creative tools) or behavior on Snapchat — and to all Snapchatters." The political affiliation of users is not assessed at any stage of our content moderation or policy enforcement processes.

16. Do you condemn Hamas' terrorist attacks on the State of Israel on October 7, 2023?

Yes.

17. What role do you believe social media companies have in promoting or limiting public speech regarding the events of October 7, 2023?

We have a zero-tolerance policy for hate speech and discrimination, and we condemn content that encourages dangerous and illegal behavior, like encouraging violence. We deliberately designed Snapchat differently than traditional social media services and don't facilitate unvetted content going viral or being algorithmically promoted to large audiences. Instead, we vet content before it can be recommended to a large audience, which helps protect against the distribution of potentially harmful or dangerous content.

In response to the events of October 7, 2023, we mobilized a dedicated cross functional team, including Trust & Safety, Content Moderation, Engineering, Policy, Legal, and Global Security, to take action on content that violates our Community Guidelines. These guidelines apply to all content, and prohibit malicious disinformation, hate speech, terrorism, violence (including graphic violence) and violent extremism.

18. What investments has your company made in anti-CSAM technology?

The sexual exploitation of any young person is horrific, illegal and against our policies. We prevent the distribution of CSAM in three key ways:

- We use cutting edge technology – PhotoDNA and CSAI Match – to proactively identify known CSAM photos and videos uploaded to Snapchat and report them to NCMEC.
- When we enforce against illegal and abusive content, we retain it for an extended period so that it is available if law enforcement want to follow up with a valid legal request.
- We take urgent action when we receive reports of CSAM, generally responding within 15 minutes.

We are actively exploring multiple privacy-respective ways of utilizing technology like Google's Content Safety API to detect novel CSEAI. We are exploring our own approaches as well, partnering with experts and organizations to potentially train our own bespoke image classifiers for photos, videos, and synthetic imagery. We will continue to iterate with our industry peers and expert organizations on ways to address these challenges.

19. Have you read the Fifth Circuit's opinion in Missouri v. Biden, No. 23-30445?

Yes.

20. Do you dispute any factual findings in the Fifth Circuit's opinions or the district court's opinions?

Neither the Fifth Circuit nor the district court opinions had any factual findings concerning Snap or Snapchat. Snap is not in a position to dispute any of the factual findings with respect to the interactions between the federal government and other online services.

21. Does your platform continue to receive requests from federal agencies to censor or promote certain content?

No, we have not received requests from federal agencies to censor or promote particular types of user content. We apply our Community Guidelines consistently, and they are publicly available here.

22. What steps do your platforms take to verify and enforce age restrictions?

When registering a Snapchat account, users are required to accept Snap's Terms of Service, acknowledge the Privacy Policy, and provide their date of birth. Users need to be at least 13 to create a Snapchat account, and the registration process fails if a user is under the age of 13. If we become aware that a Snapchat user is under the age of 13, we will terminate that user's account and delete the user's data. We also implement a safeguard that prevents younger Snapchat users with existing accounts from updating their birthday to an age of 18 or above. Specifically, if a Snapchat user between the age of 13 to 17 attempts to update their year of birth to reach an age over 18, we will prevent the change.

23. In cases where a child's safety is at risk, how does your company collaborate with law enforcement? What information or assistance is provided?

Supporting law enforcement in a timely manner is a priority for our Law Enforcement Operations team, which responds to most legal processes within two to three weeks. In the event we receive an emergency disclosure request, involving an imminent danger of death or serious injury, our team usually responds within 30 minutes. In addition to supporting all valid requests from law enforcement, our team works to proactively and quickly escalate to law enforcement any content involving imminent threats to life. For example, in 2023, we escalated more than 3,000 imminent threat situations to law enforcement.

24. Do you believe there is any expressive value in CGI or AI generated CSAM?

No.

25. Do you believe that CGI or AI generated CSAM is protected by the First Amendment to the U.S. Constitution?

Our tools are designed to prevent people from creating CGI or AI generated CSAM. We have not analyzed whether such material is protected by the First Amendment.

26. What measures does your platform take to ensure that children only see age-appropriate advertisements?

In accordance with the law, we limit distribution of certain ads to certain groups (e.g. only targeting alcohol ads to 21+ in the US). For Snapchatters aged 13 to 17, we do NOT allow ads for things like prescription medication, gambling, alcohol, tobacco, sexualized content, weight loss products or dating apps. All ads on Snapchat must adhere to our Advertising Policies, Community Guidelines and Terms of Service. We have a moderation team that reviews ads for compliance, including ads that are reported by our community.

27. Will you commit to setting up a compensation fund for those who have been harmed by your platform?

We will continue to invest in trust and safety initiatives and work with law enforcement to hold bad actors accountable for their actions. We're exceedingly grateful to those who, under horrific circumstances, have shared their experiences with us, and those stories drive us to continuously improve. The steps we've taken as a company can't reverse the tragedies they've experienced, but we all have the same goal – to keep young people – and, indeed, all people – safe when using our service.

28. How many sales of fentanyl are transacted through Snap each year?

No fentanyl sales are transacted through Snapchat. There is no user-to-user marketplace feature on Snapchat, and Snapchat does not have a system for payments between Snapchat users. We work to prevent users from communicating about drug sales on Snapchat by proactively detecting drug-related content and blocking dealers from using our service.

This conduct is illegal and we invest a significant amount of resources in attempting to make Snapchat an inhospitable service for drug dealers to communicate about drug deals.

29. What steps is Snap taking to eradicate traffic of illegal narcotics on the platform?

Snapchat has been working for years to remove drug dealers from our service. We block drug-related search terms and respond to those queries with educational content. We proactively detect and remove drug-related content including powders and pills. When we enforce against illegal and abusive content, we retain it for an extended period so that it is available if law enforcement wants to follow up with a valid legal request, and we also make proactive referrals for prosecution. We've collaborated with other services and with NGOs to support some of the country's most significant public education campaigns so that people know one pill can kill. We launched signal sharing with Meta because we know criminals use multiple services and sharing

signals can help us keep people safe. We also support a legislative approach, including the Cooper Davis Act, which we have supported since its inception.

30. What steps is Snap taking to eradicate child pornography on the platform?

As to Snap's efforts to combat child pornography specifically, please see response to question 18.

31. Why did Snap seek sanctions in case No. 3:22-cv-619 in the Southern District of California against plaintiffs who alleged that they were underage victims contacted by adult male perpetrators and groomed on your platform?

Snap sought sanctions against the attorneys who filed that case because the complaint contains several clear falsehoods about how Snapchat works. For example, the complaint falsely alleged that Snapchat's friend suggestion feature, Quick Add, would suggest friends based solely on geographic proximity or shared interests, trying to make the Court believe that a child predator could get friend recommendations for minors on Snapchat simply by loitering near a school or feigning an interest in topics that appeal to minors. That is not how Quick Add works. Quick Add does not recommend users based merely on shared interests or because users may be near a similar location like a school or a park. Quick Add instead primarily makes friend suggestions based on a user's social graph, such as having mutual friends in common. The purpose of QuickAdd is to connect a user with other Snapchat users they already know, not strangers.

The complaint contained many other demonstrably false accusations, as well. It is sanctionable conduct for attorneys to put known falsehoods in their complaints in an attempt to survive a pleading challenge. Because Snap is not able to contest the factual allegations in Plaintiffs' complaint on a motion to dismiss, Snap had to file the motion for sanctions seeking to have the false paragraphs removed from the complaint to correct the record. The court concluded that the Plaintiffs' allegations were "tenuous and confounding," that other allegations were "a stretch, suggestive of sloppy drafting," and that some of the allegations "lack apparent basis in fact."

32. Does Snap have access to and provide copies of users' direct message communications to federal law enforcement upon request?

Upon receipt of valid legal process which authorizes such a disclosure, Snap will provide all communications content in our possession for the implicated account.

33. How does Snap test new products (such as My AI) to determine whether they are safe and healthy for children?

We prioritize the safety and wellbeing of our community, making it a core aspect of our product development. Snapchat features, especially those incorporating AI, undergo a meticulous review process. This involves a rigorous evaluation based on our Safety and Privacy by Design

principles, ensuring that aligns with our commitment to user safety. MyAI is powered by OpenAI's technology, and we've added extra safeguards to help our community have a positive and age appropriate experience.

Before a person can use My AI, they receive a message from us making clear it's a chatbot, that it has limitations, and that we treat the data in these messages differently than conversations with human friends – this helps us strengthen the overall experience, make improvements when the chatbot responds incorrectly, and enhances our protections. We programmed My AI to abide by our Community Guidelines, and to consider a Snapchatter's age group in its responses. We pause a person's access to My AI if they repeatedly try to misuse it.

We also have rolled out more controls for parents through our Family Center tool. Through Family Center, parents have visibility into when their teen is chatting with My AI, and have the option to restrict their teen's access to My AI.

We regularly analyze My AI's responses so we can keep improving it – and continue to find that only a small percentage (0.01%) are "non-conforming," meaning they don't adhere to our policies. A common example of this is when a user asks a question that includes an inappropriate word (sometimes in an attempt to "trick" My AI) and My AI repeats that inappropriate word in response.

34. Has Snap conducted internal studies on the psychological effects of filters on users' self-esteem and body image? If not, why not?

Snap has conducted research on filters relating to a range of topics. For example, Snap has conducted research into how users engage with lenses and, in the course of those studies, participants provide a range of responses describing what they enjoy and don't enjoy about those lenses. Snap then uses the collective feedback to inform product development.

35. What efforts has Snap made to educate users about the potential impact of excessive filter usage on body image?

It is worth noting that the majority of lenses on Snapchat are intended to be lighthearted and silly filters where users can turn themselves into a hot dog, Halloween character, or other fun creative persona. And when someone elects to use a lens, there is an indicator to other Snapchatters that the image was created with a lens, empowering the recipient to understand how the visual effects have been achieved.

As part of our overall effort to prioritize the mental health of Snapchatters, in 2020, we launched Here For You, a proactive in-app support system for Snapchatters focused on mental health and well-being. Here for You surfaces resources from expert organizations when Snapchatters search on a range of mental health-related topics, including eating disorders, anxiety, stress, depression,

hate speech, bullying, suicidal thoughts, and more. We partner with the National Eating Disorders Association (NEDA) and the National Alliance for Eating Disorders (The Alliance), whose resources appear in the tool.

36. Does Snap monitor user behavior to identify potential signs of negative psychological effects related to filter use?

We have Organic and Community AR Guidelines for accepting Lenses into the app. While they do not identify rejecting Lenses for potential signs of negative psychological harm, we do highlight the following policies for Organic Lenses:

- We design every Lens with race, gender, ethnicity, and cultural norms in mind. We leverage our ever-growing diverse training datasets, as well as feedback from community members to do this.
- If a Lens does not resonate with our community, as expressed through a high ratio of user reports, we take that feedback into consideration and will re-review the Lens with a goal to leave as-is, modify, or remove.
- We consider current and historical global events when releasing a Lens, and delay or deny amplification to Lenses that may be deemed insensitive due to broader social occurrences throughout the world
- Lenses should not change your skin tone to mimic a different ethnicity or race
- We do not modify facial or other features in a way that evoke racial, ethnic, cultural or religious stereotypes or stigmatized disabilities
- We present religious and cultural iconography in a respectful manner, with feedback solicited from internal and external subject matter experts. This means we are especially thoughtful around holiday or event-based content, including the geography in which a Les will launch.
- We ensure that a Lens is not deceptive. We use signifiers and watermarks where there may be questions of creative authenticity
- We test Lenses on photos/videos of and in real life settings with diverse groups of people to accurately enforce our policies

Community Lenses submitted by our users are monitored, and rejected, for the following reasons related to potential impact on users:

- Body shape or size
    - Proportion changes that emphasize sexualized body parts
    - "Fat-shaming"
    - Realistic imitations of eating disorders
- Facial Features
    - Face swaps ("deep fakes") that could be mistaken as authentic (because of seamless quality without a signifier or watermark)

- Intentional racial or ethnic stereotypes
- Mimicry of real-life medical conditions, disabilities, stigmatized ailments, or specifically imitating the effects of an eating disorder

When our Trust & Safety team recognizes a Snapchatter in distress, they will forward self-harm prevention and support resources, and notify emergency response personnel as appropriate. The resources that we share are available on our global list of safety resources, and these are publicly available to all Snapchatters. We have also developed a Safety and Privacy Hub that provides additional details on our Safety and Privacy by design approach, as well as resources for reporting safety concerns and guidance for staying safe on Snapchat.

37. When users exhibit signs of body dysmorphia, eating disorders, or related issues, what measures does Snap have in place to provide support or resources?

We integrate safeguards and tools we use across Snapchat, including blocking results for many harmful topics and providing in-app resources which show Snapchatters resources from expert partners when they search for content related to eating disorders, mental health, anxiety, depression, bullying and related topics. Our Community Guidelines strictly prohibit the glorification of eating disorders and self-harm, gratuitous violence, bullying and harassment, or sending a Snap with the intention of hurting or harming another person.

As part of our overall effort to prioritize the mental health of Snapchatters, in 2020, we launched Here For You, a proactive in-app support system for Snapchatters who may be experiencing a mental health or emotional crisis by surfacing resources from expert organizations when Snapchatters search on a range of mental health-related topics, including eating disorders, anxiety, stress, depression, hate speech, bullying, suicidal thoughts, and more. We partner with the National Eating Disorders Association (NEDA) and the National Alliance for Eating Disorders (The Alliance), whose resources appear in the tool.

When our Trust & Safety team recognizes a Snapchatter in distress, they will forward self-harm prevention and support resources, and notify emergency response personnel as appropriate. The resources that we share are available on our global list of safety resources, and these are publicly available to all Snapchatters. We have also developed a Safety and Privacy Hub that provides additional details on our Safety and Privacy by design approach, as well as resources for reporting safety concerns and guidance for staying safe on Snapchat.

38. What is the total dollar value of all drug deals facilitated by Snap since its inception?

Snapchat does not facilitate drug deals. There is no user-to-user marketplace feature on Snapchat, and Snapchat does not have a system for payments between Snapchat users. We proactively detect and remove drug-related content and block dealers from using our service.

Sen. Mike Lee (R-UT)

1. The 2022 Thorn Report identified Snapchat as the #1 platform where most minors (21 percent) reported having an online sexual encounter with someone they believed to be an adult (14 percent). Amongst minors who share self-generated CSAM, 52 percent say they do so with people they only know from online interactions. What is Snapchat doing to cease being the preferred platform for predators to sexually exploit children?

We know that Snapchat is used by the vast majority of American teenagers, which we believe contributes to the prevalence of these interactions when compared to other services which are not used as frequently by teenagers. Accordingly, we have an enormous responsibility to help keep our community safe. We've designed Snapchat to make it difficult for predators to find and interact with teens.

There are no public friend lists, and by default teens must have proactively added someone as a friend or have them in their contact book to receive a direct message from a user. That's different from text messaging where anyone who has your phone number can get in touch. We also have extra safeguards to block predators from being able to find and search for teens, and to warn teens if an adult tries to contact them who isn't a mutual friend or existing phone contact.

In addition, we are investing resources in learning about the techniques that these predators use, so that we can identify possible sexually exploitative conduct and quickly take action against the accounts. When we disable accounts for sexual exploitation and grooming behavior, we also take steps to block the associated device from creating another account on our platform. Predators' techniques are always evolving, which requires continued focus on the part of all platforms.

We also offer easy reporting mechanisms so teens can get help quickly, and we typically respond to these reports in under an hour.

2. Snapchat restricts certain content from accounts that belong to minors. However, the only age verification measure that Snapchat undertakes to ascertain the age of its users is asking new users to enter their birthdate when they open an account. How do you prevent minors from lying about their age when creating an account?

We cannot prevent minors from lying about their age. We do not want people under the age of thirteen to use Snapchat, and if we determine that an account is used by someone under thirteen we remove the account from our service. We strongly recommend that parents or caregivers who provide a smartphone to children under the age of thirteen utilize the operating system-level parental controls to set the child's age accurately and restrict the downloading of apps that are not intended for children.

3. Digital sextortion is a growing epidemic on all social media platforms. The majority of victims of sextortion were approached by predators on your platform. You have increased the number of employees tasked with child safety, and you automatically report suspected grooming to NCMEC. How do you ensure that these predators cease contacting minors? How do you catch these situations before a predator attempts to get the minor to move to another platform?

We are very much aware and focused on the growing rise in financially motivated "sextortion" — where criminals pose as young people and trick victims into sending compromising images. This conduct is illegal and abhorrent, and we take active steps to identify and prevent such exploitation as well as empowering users to identify and report suspicious contact.

We combat this criminal activity by:

- Making it difficult for strangers to search for, find, or contact teens.
- Warning teens if they receive a friend request from another person who isn't a mutual friend or phone contact.
- Keeping friend lists private and preventing teens from having public profiles — which helps prevent criminals from using Snapchat to target a teen's friends
- We proactively detect accounts attempting to engage in sextortion and disable them. When we remove sextortion-related content, we also retain it for an extended period of time in case law enforcement wants to follow up with a valid legal request.
- We've added an in-app reporting option specifically for sextortion ("They leaked/are threatening to leak my nudes.") to make it easier for our community to report such abuse to us. When we receive such a report, we take action quickly – usually within 30 minutes.
- We also educate teens about the dangers of this type of crime and urge them to use our in-app reporting tools.

This is very important because young people may feel afraid to report the problem to their parents or caregivers.

4. How does Snapchat inform parents when a child is exposed to sexual material? How does Snapchat inform parents when their child is the target of grooming?

Snap's parental resource, Family Center, is designed to spark meaningful conversations between parents and teens about who the teen is communicating with on Snapchat and the content the teen is viewing. For parents and teens who are part of Family Center, parents can see, among other things, who their teen is friends with and who they have been communicating with over the last week, without compromising the teen's privacy by disclosing the content of any messages. The parent can also set content controls. If a parent is concerned about a particular friend, the parent can easily report the account to Snapchat for review. At its core, Family Center seeks to

spark dialogue between teens and their caregivers about staying safe on the app and online generally.

5. Besides scanning uploaded pictures for known CSAM, what other measures is Snapchat taking to prevent minors from sharing self-generated CSAM with others?

Snap engages in awareness-raising and education about the consequences of creating self-generated CSAM. We make available in-app a channel called "Safety Snapshot," which includes an episode about sexting and sharing self-generated CSAM. The episode seeks to speak to teens in a relatable language and style, and offers links to helpful resources. It was reviewed and approved by experts at the National Center for Missing and Exploited Children (NCMEC) prior to release.

In addition, we participate in NCMEC's Take It Down program, which allows minors to generate a digital fingerprint (called a "hash") of selected images(s) and video(s) directly on their device (i.e., cellphone, computer, tablet). Participating companies, including Snap, can then use those hashes to detect matches and remove imagery that violates our Community Guidelines. We help to evangelize the availability of the Take It Down program to our community in communications with under-18 sextortion victims, encouraging them to leverage the service. We will continue to iterate with our industry peers and expert organizations on ways to address these challenges.

We are also actively exploring multiple privacy-protecting ways of utilizing technology like Google's Content Safety API to detect novel CSAM. We are exploring our own approaches as well by partnering with experts and organizations to potentially train our own bespoke image classifiers for photos, videos, and synthetic imagery. We will continue to iterate with our industry peers and expert organizations on ways to address these challenges.

6. Snapchat does not monitor the content of conversations between users, which contributes to rampant illegal activity on your platform. This includes drug trafficking, weapons trafficking, and child sexual exploitation. What are you doing to eliminate these types of exchanges? Should you do more?

Snap invests a significant amount in safety and combating harms, and more than 20% of our team works on safety-related matters. For example, Snapchat has been working for years to remove drug dealers from our service. We block drug-related search terms and respond to those queries with educational content. We proactively detect and remove drug-related content including powders and pills. When we enforce against illegal and abusive content, we retain it for an extended period so that it is available if law enforcement wants to follow up with a valid legal request, and we also make proactive referrals for prosecution. We've collaborated with other services and with NGOs to support some of the country's most significant public education campaigns so that people know one pill can kill. We launched signal sharing with Meta because we know criminals use multiple services and sharing signals can help us keep people safe.

We support a legislative approach to helping combat drug activity online, including the Cooper Davis Act, which we have supported since its inception. Additionally we are proud to support the Kids Online Safety Act, which will create standards for all services to protect the privacy and safety of young people online and have already implemented many of its provisions. We are happy to provide additional detail on how we address other severe harms as well.

Snap is aware of no communications service that proactively monitors all of the content of private conversations between users. Telephone service providers do not listen to their customers' conversations looking for evidence of illegality. Email service providers do not read their customers' emails looking for evidence of illegality. The postal service does not open and read letters transmitted by mail looking for evidence of illegality. For Snap to screen private messages looking for evidence of illegality would be a gross invasion of our community's privacy, and we do not believe such monitoring is warranted when the vast majority of our community uses Snapchat to communicate safely with their friends and family.

7. Do you intend on employing end-to-end encryption for accounts that belong to minors?

We believe encryption is important in helping people communicate with their friends and family privately and safely, but we do not plan to implement end-to-end encryption in a way that would prevent us from being able to detect known child sexual abuse material or constrain our ability to report such content to the authorities.

8. Snapchat Stories are a growing concern among parents, and for good reason. From the inception of that feature, those public stories have been plagued by sexually explicit, degrading, and morally repugnant material. This issue was brought to your attention by the Senate Commerce Committee in October, 2021, and despite promises of increased moderation by Snap, Snapchat Stories continue to host reprehensive material. Why are creators using Snapchat Stories permitted to upload this type of content for public consumption?

We work to ensure that Snapchatters have an age appropriate experience across all surfaces on our app. On Snapchat's public content surfaces — which includes Spotlight and Snapchat Stories (also known as Discover) – we do this in a few ways:

- We moderate content before it can be recommended to a large audience; all content is assessed against our Content Guidelines for Recommendation Eligibility. We have introduced new moderation techniques, as well as safeguards to ensure teen users are age-gated from suggestive content on Discover by default.
- We use automation (such as signal-based detection, machine learning, and keyword lists) to proactively identify and remove certain types of harmful or illegal content upon submission (including CSAM, illicit-drug, and sextortion-related content).
- We work with NCMEC, law enforcement, and industry partners to identify, remove, and escalate harmful content and to remain vigilant against emerging risks and trends.

- For infractions that are less egregious, we use a Strike System to crack down on accounts that repeatedly attempt to share content that violates our Community Guidelines.

We also know parents may have different comfort levels on what types of content their teen can view based on their maturity and family values. Our Family Center tools allow parents to set controls that filter out suggestive or sensitive content from recommendations in Spotlight or Discover.

9. In January of this year, a researcher at NCOSE created a fake Snapchat account posing as a 13-year-old. Within ten minutes, that account accessed videos simulating sex acts, men slapping women's barely-clothed rears, strip club promotions, and stories about how men should sexually pleasure women—all through the Snapchat Stories section. How do you permit accounts—that you know belong to minors—to access these types of materials? Why are these stories hosted and promoted by Snapchat? What will you do to put an end to a minor's ability to access these promoted materials?

NCOSE made us aware of these findings, and we immediately investigated. We uncovered an issue in Saved Stories due to a gap in our suggestive filtering logic implementation. As of January 31, that gap was closed. It is important to note that this issue involved suggestive content, not explicit content. We detect and remove explicit content automatically through machine-learning technology. We want all Snapchat users to have an age-appropriate content experience – especially teens – and aim to provide this in a few ways:

- We moderate public content before it can be recommended to a large audience, and have strict content guidelines for what is permitted on Snapchat.
- We use machine learning to proactively find sexually explicit content and accounts, so we can remove them.
- We use a Strike System to crack down on accounts trying to market this content.

We also know parents may have different comfort levels as to the types of content they would like their teens to view based on the teen's age, their maturity level, and their family values. Our Family Center tools offer parents the ability to set controls that filter out even further suggestive or sensitive content. As of 2023, new joiners to Family Center have these Content Controls turned "on" by default.

10. The Department of Justice reported that Snapchat's "Suggested Friends" feature provides an avenue for a predator to easily access an entire network of minors. Once one minor accepts the request from a predator, that predator can easily move from one minor's network to another without end. Will you discontinue the Suggested Friends feature for accounts belonging to minors?

Quick Add is a tool that makes it easier to find close friends. We make sure that a young person under 18 is not recommended to another person unless they have multiple mutual friends, meaning they are likely to know them. Quick Add does not recommend users based on shared interests or because users may be near a similar location, like a school or a park. Quick Add instead primarily makes friend suggestions based on a user's social graph, such as having mutual friends in common. The purpose of QuickAdd is to connect a user with other Snapchat users they already know, not strangers.

We also show warnings to teens if someone with whom they do not share a mutual friend tries to chat with them.

Sen. Alex Padilla (D-CA)

1. In recent years, more companies in the tech sector are offering tools to enable caregivers to have a dialogue with minors in their care about healthy and safe internet activity. An important element in understanding whether these tools are helpful is understanding whether or not these tools are being adopted. You were the only CEO prepared to answer questions about adoption rates, and you shared at the hearing that approximately 20 million teens use Snapchat in the United States, that around 200,000 parents use the Family Center supervision controls, and that 400,000 teen accounts have been linked to a parent's account through Family Center.

a. How are you ensuring that young people and their caregivers are aware of these tools?

We are constantly working to educate Snapchatters and parents about the accessibility of tools such as Family Center. We promote Family Center at large online safety conferences and events, as well as smaller gatherings and meetings. We engage with nonprofits and NGOs to help make parent groups and related organizations aware of Family Center, including with influential groups such as the National Parent Teacher Association (PTA) and the American Federation of Teachers. We also invest in search engine optimization advertising, allocating budget to Google search ads so that Family Center information is ranked at the top for parents looking for information on parental tools.

b. How are you ensuring that these tools are helpful to both minors and their caregivers?

Prior to Family Center's launch in the U.S. in August 2022, and on several occasions since, we have conducted focus groups with both parents and teens. We also sought feedback and input from more than 40 child safety experts across the globe before Family Center was released. We continue to consult our Safety Advisory Board on updates and improvements to Family Center.

This year, we are excited to launch Snap's Teen Council to hear ideas for continuing to make Snapchat a safer, healthier, and more enjoyable place for creativity and connecting with friends. We appreciate that navigating online can present risks and we want to make sure that young people understand, can recognize, and have the skills to help mitigate those risks. For any teen who participates in our Teen Council, there will also be information for their parents and guardians as well as opportunities to seek their feedback.

2. Snap offers a broad range of "user empowerment" tools, and it's helpful for policymakers to understand whether young people even find these tools helpful or are actually adopting them. Additionally, some safety features still put the onus on young people to employ a great deal of judgment about safety.

a. How are you ensuring that the burden is not on young people to make adult-level decisions about safety on the services that you operate?

Snap is committed to protecting the privacy and wellbeing of our community, which is why we were the first technology company to endorse the Kids Online Safety Act. In fact we have already implemented many of its provisions such as mandating appropriate default privacy settings, providing safeguards for teens (including additional privacy settings that protect their privacy and offer them control over their experience), offering parental tools, and limiting the collection and storage of personal information.

Protecting the privacy and safety of young people on Snapchat is a top priority, and we did not wait for this bill to be implemented to make these changes.

> **b.** Over the last 4 years, how often have you blocked products from launching because they were not safe enough for children, or withdrawn products from the market after receiving feedback on the harms they were causing?

Our Safety by Design process is structured so that cross-functional teams collaborate in identifying and addressing potential safety risks during the development of a product, often starting at early phases of ideation. Further, beyond safety, there are many considerations that go into product development and decisions about whether to roll out or test products, including technical considerations, resources, testing results, and feedback.

3. Existing detection tools for keeping child sexual abuse material from spreading online rely on hashed images of already identified CSAM imagery. There are tools like PhotoDNA and Google's CSAI match tool available for identifying this content. A challenge I hear raised frequently is identifying and removing novel images that have not already been hashed.

> **a.** What would it take to develop better technology to accurately identify and limit the spread of novel CSAM images?

The issue of novel CSAM is a complicated one, but we are actively exploring multiple privacy-protecting ways of utilizing technology like Google's Content Safety API to detect novel CSEAI. We are exploring our own approaches as well, partnering with experts and organizations to potentially train our own bespoke image classifiers for photos, videos, and synthetic imagery. Last year, we joined "Take It Down," NCMEC's program to help remove online nude, partially nude, or sexually explicit photos and videos of minors. We also started evangelizing the existence of, and our participation in, Take It Down in communications with under-18 sextortion victims, encouraging them to leverage the service. We will continue to iterate with our industry peers and expert organizations on ways to address these challenges.

We believe that establishing a legal safe harbor for companies to use CSAM content reported to NCMEC to train machine learning models for novel-CSAM detection is critical. Our team

would be very happy to work with your office — in collaboration with law enforcement and other child safety experts — to help develop a responsible approach to such a framework.

b. Are there interventions from Congress that would facilitate identification of CSAM?

Machine-learning tools offer one method by which novel CSAM might be detected. However, companies are barred by current law from using existing CSAM that has been reported to NCMEC in the dataset used to train these tools. Intervention by Congress to permit such use by researchers could help facilitate the creation of effective machine-learning tools for identifying novel CSAM. Our team would be very happy to work with your office — in collaboration with law enforcement and other child safety experts — to help develop a responsible approach to such a framework.

c. Based on your company's experience trying to address online sexual exploitation and abuse of minors, are there areas where Congress could be helpful in tackling this problem?

The creation and spread of AI-generated CSAM is quickly evolving and we welcome the collaboration of our industry partners, the government and law enforcement to combat this threat. Congress could help by providing a responsible framework that allows companies to legally and safely test large language models to help detect AI-generated CSAM. Our team would be very happy to work with your office, child safety, experts and law enforcement on how best to do that.

4. AI models are making it easier to develop synthetic CSAM. These are either altered images of real people, or wholly synthetic individuals. Policymakers are grappling with what this will mean for law enforcement efforts to hold perpetrators accountable and identify children who are being harmed. In addition to processing a higher volume of Cybertips, investigators will have the added challenge of determining whether the victim in the scenario is in fact a real person. And cases are already being reported where AI generative technologies are being employed to facilitate the grooming and sextortion of minor victims.

a. What are you doing to identify and remove AI-generated CSAM on your services?

We prohibit all CSAM on Snapchat, including content that is AI-generated, and use technology safeguards to help ensure this content is not created on Snapchat and, if we identify it, remove it promptly and report it to NCMEC as appropriate.

The creation and spread of AI-generated CSAM is quickly evolving and we welcome the collaboration of our industry partners, the government and law enforcement to combat this threat. One important way we can all get better at preparing for it is to have a responsible framework that allows companies to legally and safely test large language models.

b. Do you flag for NCMEC if you perceive the CSAM to be AI-generated?

Yes.

c. How prevalent is this kind of content?

We understand the creation and spread of AI-generated CSAM is quickly evolving but Snap has seen very little of this content to date.

d. How do you anticipate the rise of AI-generated CSAM will impact NCMEC's ability to process and refer Cybertips to law enforcement?

AI-generated CSEAI, unfortunately, is already proving challenging for NCMEC and hotlines across the globe. Hotline analysts need to closely analyze photos for "tells" (e.g., extra digits, unnatural body positioning) as to whether the images may be wholly AI-generated or include partial depictions of real children. Either way, the process will take longer and delay referrals to law enforcement, as well as take away time from vital review cycles of real illegal imagery.

e. Recently, A.I.–Generated explicit images of a major pop superstar were distributed widely online without her consent. That story drew attention to a growing problem over the last year facilitated by AI tools: the generation of deepfake, nonconsensual, sexually explicit imagery of everyday people, including our young people. Will you commit to reporting on the prevalence of this new problem and the steps your company is taking to address this horrendous abuse?

We are committed to reporting suspected AI-Gen CSEAI to NCMEC and are committed to utilizing reliable tools that have been developed to identify it.

f. Are there technical or legal barriers that your company has identified preventing thorough redteaming of AI models to ensure they do not generate CSAM?

Evaluating whether models are capable of producing CSEAI is important, yet there is complexity and ambiguity in the legal landscape that makes that difficult. We believe that establishing a legal safe harbor for controlled testing and red-teaming in this area would help to advance child safety, and our team would be very happy to work with your offices — in collaboration with law enforcement and other child safety experts — to help develop a responsible approach to such a framework.

5. How companies choose to allocate their resources illustrates their true priorities.

a. What percentage of your company's budget is dedicated to addressing child safety on your platform?

Approximately 10% of Snap's overall personnel expense goes toward the core teams that work on safety and content moderation.

b. What process or assessment of risk on the platform informed that figure?

Safety is a top priority for Snap. We are constantly evaluating safety risks and reassessing resourcing to ensure that safety issues are appropriately prioritized. More specifically, we take into consideration a combination of the results of our twice annual voluntary Transparency Reports, our internal Harms Prioritization Framework, platform safety metrics and an ongoing survey of the threat landscape to ensure we are dedicating sufficient resources to the most serious risks and potential harms.

c. How many layers of leadership separates your trust and safety leaders from you?

Our Chief Security Officer, who reports to me, supervises the Senior Director who oversees the Trust & Safety team and Law Enforcement Operations teams. I also meet directly with the leaders of the Trust & Safety team regularly to review our safety roadmap.

6. The companies represented at the hearing have the money and resources to hire teams of Trust & Safety professionals and build bespoke tools to aid with content moderation and integrity work as well as the detection of content like CSAM on their services. This is not necessarily the case for the rest of the tech sector. These are industry-wide problems and will demand industry-wide professionalization and work.

a. What is Snap currently doing to support access to open-source trust & safety tools for the broader tech ecosystem?

We work with expert NGOs, the government, and our industry peers to collectively attack this problem, including supporting efforts like NCMEC's "Take It Down" program. Take It Down helps to remove online nude, partially nude, or sexually explicit photos and videos of minors. We also started evangelizing the existence of, and our participation in, Take It Down in communications with under-18 sextortion victims, encouraging them to leverage the service.

In addition, Snap is an active member of the Technology Coalition, a group of 37 tech companies working to end online child sexual exploitation and abuse. Snap has chaired a number of Tech Coalition working groups, including the groups on Tech Innovation, Transparency and Accountability, and Collective Action.

b. And if Snap is not doing anything now, will you commit to supporting the development of these kinds of resources?

N/A

7. One necessary element of keeping our kids safe is preventing harm in the first place. The National Center for Missing and Exploited Children partnered with the White House, the Department of Justice, and the Department of Homeland Security to create "The Safety Pledge" initiative to combat online child exploitation in September 2020. I understand more government backed public awareness campaigns are being developed.

a. Are you partnering with the federal government to distribute health and safety resources to young people?

Yes. Snap was the first private sector company to sign a Memorandum of Understanding with the Department of Homeland Security in support of DHS's upcoming Know2Protect (K2P) campaign. K2P is designed to raise awareness of the risks of child sexual exploitation and abuse online. Snap will provide an ongoing ad grant to DHS for free advertising on the service, and support and promote the campaign among the Snapchat community, parents, and NGOs in a number of ways. Snap is a strong proponent of public-private partnerships that offer a single, galvanizing message and program to raise awareness and educate users about key safety issues.

b. What are you proactively doing to educate the minors that use your services about online health and safety?

As part of our overall effort to prioritize the mental health of Snapchatters, in 2020, we launched Here For You, a proactive in-app support system for Snapchatters who may be experiencing a mental health or emotional crisis by surfacing resources from expert organizations when Snapchatters search on a range of mental health-related topics, including eating disorders, anxiety, stress, depression, hate speech, bullying, suicidal thoughts, and more. When our Trust & Safety team recognizes a Snapchatter in distress, they will forward self-harm prevention and support resources, and notify emergency response personnel as appropriate. The resources that we share are available on our global list of safety resources, and these are publicly available to all Snapchatters.

In 2023, we launched four new "Safety Snapshot" episodes in-app to help raise awareness and educate our community about various online sexual-related risks, namely: sexting and nudes, financial sextortion, child sex trafficking, and child online grooming for sexual purposes. Safety Snapshot is our official in-app channel for online safety and privacy advice and guidance. These episodes seek to speak to teens in a relatable language and style, and offer links to helpful resources. All four of the episodes were reviewed by experts at the National Center for Missing

and Exploited Children (NCMEC) prior to release.

We have also provided users on Snapchat with safety messaging regarding the risks of fentanyl poisoning and counterfeit prescription drugs. We developed and made available an in-app education portal called "Heads Up" that distributes content from expert organizations such as Song for Charlie, Shatterproof, the CDC and the Substances and Mental Health Services Administration. If someone on Snapchat searches for drug-related keywords, Heads Up will show relevant educational content designed to dissuade engagement and ultimately prevent harm to our community.

Additionally, in partnership with Song for Charlie, we developed a video advertising campaign that has been viewed over 260 million times on Snapchat, and rolled out a new national filter that raises awareness of the dangers of fentanyl and counterfeit pills and directs Snapchatters to the Heads Up educational portal.

Finally, Good Luck America, a Snap Original news show, produced a special edition series of episodes devoted to educating our community about the fentanyl crisis.

8. Sextortion has become increasingly prevalent. Offenders may use grooming techniques or basic trickery to manipulate victims into providing nude or partially nude images of themselves, which are then used to coerce victims into sending more graphic images and videos or pay a ransom. These criminals often threaten to post the images or sensitive images publicly or send them to the victim's friends and family if the child does not comply. From May 2022 to October 2022, U.S. law enforcement and NCMEC witnessed an alarming increase in CyberTips and reports where minors have been sextorted for money. Many young boys, including in California, have committed suicide out of desperation, leaving their loved ones devastated.

a. How is your company responding to the growing threat of financial sextortion?

This conduct is illegal and abhorrent, and we take active steps to identify and prevent such exploitation as well as empowering users to identify and report suspicious contact.

We combat this criminal activity by:

- Making it difficult for strangers to search for, find, or contact teens.
- Warning teens if they receive a friend request from another person who isn't a mutual friend or phone contact.
- Keeping friend lists private and preventing teens from having public profiles — which helps prevent criminals from using Snapchat to target a teen's friends.
- We proactively detect accounts attempting to engage in sextortion and disable them. When we remove sextortion-related content, we also retain it for an extended period of time in case law enforcement wants to follow up with a valid legal request.

- We've added an in-app reporting option specifically for sextortion ("They leaked/are threatening to leak my nudes.") to make it easier for our community to report such abuse to us. When we receive such a report, we take action quickly – usually within 30 minutes.
- We also educate teens about the dangers of this type of crime and urge them to use our in-app reporting tools.

This is very important because young people may feel afraid to report the problem to their parents or caregivers.

b. What methods are in place to detect and disrupt this type of abuse in real time?

Please see response to question 8.a.

c. What kind of user education and awareness are you engaged in?

In 2023, we launched four new "Safety Snapshot" episodes in-app to help raise awareness and educate our community about various online sexual-related risks, namely: sexting and nudes, financial sextortion, child sex trafficking, and child online grooming for sexual purposes. Safety Snapshot is our official in-app channel for online safety and privacy advice and guidance. These episodes seek to speak to teens in a relatable language and style, and offer links to helpful resources. All four of the episodes were reviewed by experts at the National Center for Missing and Exploited Children (NCMEC) prior to release.

d. Are you aware of a higher prevalence of sexual extortion or abuse against certain demographics among young users? If not, will you commit to studying this issue and making that kind of information available to improve public education and protection measures?

We have heard from NCMEC and other international safety hotlines, as well as anecdotally, that teen males are often targeted in sextortion schemes. That gender assessment bears out to a degree in our own research. We conducted a deeper dive into sextortion among teens (13-17) and young adults (18-24) as part of our 2023 annual Digital Well-Being Index research that we lead in six counties (Australia, Germany, France, India, UK and US). We asked users about their online risk exposure, among other things, and we asked them about their experiences on all services and devices, not just Snapchat. In that research, among those who reported being catfished and /or hacked, the gender split was 56% male and 44% female.

9. Young people need to be at the center of regulatory discussions, and they need to be at the table as products and services they use are designed.

a. Are you engaging young adults and youth in your conversations and policies around Trust and Safety on the platform?

Yes. In addition to our Safety Advisory Board, which includes three Generation Z young adults, who are also youth advocates for online safety, at the start of the year we opened applications for Snap's first Council for Digital Well-Being, a pilot program in the U.S. for 13-to-16-year-olds interested in collaborating with Snap to create an even safer and healthier environment for creativity and connection among real friends and family. We plan to select a diverse group of about 15 young people from across the nation for an 18-month-program that will include monthly calls, project work, and engagement with our global Safety Advisory Board. In this first year, selected council members will be invited to Snap's headquarters in Santa Monica, California, for a two-day summit and, in Year Two, we have plans for a more public-facing event featuring council members and showcasing their knowledge and learning. The program will also include a "parent track" for parents, guardians, and chaperones, accompanying the teens to the events.

b. How do you proactively keep up to speed with the most pressing issues facing young people online?

Our Safety Advisory Board consists of 18 members, based in 10 countries and representing 11 different geographies and regions. The Board is made up of 15 professionals from traditional online safety-focused non-profits and related organizations, as well as technologists, academics, researchers, and survivors of online harms. These members are experts in combating significant online safety risks, like child sexual exploitation and abuse and lethal drugs, and have broad experience across a range of safety-related disciplines. In addition, we have 3 Board members who are young adults and youth advocates. We selected these members to ensure the Board has ready-access to the all-important "youth voice" and viewpoint; to make certain a portion of the Board includes committed Snapchat users; and to seek to balance professional views with practical perspectives from a core demographic of the Snapchat community.

10. For many children, an open dialogue about their internet habits is a best practice, and healthy. But not every child has a parent or a caregiver that is looking out for their best interest. For many kids who are abused, a caregiver or parent is their abuser. Additionally, for many young people, their parents' knowledge of their sexual orientation or their interest in exploring it, fundamentally puts them in jeopardy. Solving for these different needs across our young people at the scale of social media and internet applications is really vital.

a. How have you designed your parental tools with this dynamic in mind?

Family Center helps parents get more insight into who their teens are friends with on Snapchat, and who they have been communicating with, while still respecting their teens' privacy and autonomy. It's designed to reflect the way parents engage with their teens in the real world, where parents usually know who their teens are friends with and when they are hanging out, but don't eavesdrop on their private conversations.

Sen. Thom Tillis (R-NC)

1. Twenty-one is the minimum age to purchase highly regulated adult products such as alcohol, tobacco, and nicotine. Nevertheless, there is a proliferation of user-generated content posted on social media sites featuring underage use of these products. Recently, some have proposed banning these age-restricted products due in part to the user- generated content being available on your respective platforms. Surely, banning these products cannot be the answer. However, we must do more – your company must do more – to shield underage audiences from exposure to this content.

Therefore, as the content moderator of these platforms, what policies do you have in place, and what more can you do, to prevent this type of user-generated content from reaching underage audiences? How do you respond to requests to pull this content from your sites when deemed inappropriate for underage audiences?

We want all Snapchat users to have an age appropriate experience – especially teens – and aim to provide this in a few ways:

- We moderate public content before it can be recommended to a large audience, and have strict content guidelines for what is permitted on Snapchat
- We use machine learning to proactively find sexually explicit content and accounts so we can remove them
- We use a Strike System to crack down on accounts trying to market this content

We also know parents may have different comfort levels on what types of content their teen can view based on their maturity and family values. Our Family Center tools allow parents to set controls that filter out suggestive or sensitive content.

In addition, in accordance with the law, we do limit distribution of certain ads to certain groups to comply with applicable law (e.g. only targeting alcohol ads to 21+ in the US). For Snapchatters aged 13 to 17, we do NOT allow ads for things like prescription medication, gambling, alcohol, tobacco, sexualized content, weight loss products or dating apps. All ads on Snapchat must adhere to our Advertising Policies, Community Guidelines and Terms of Service. We have a moderation team that reviews ads for compliance, including ads that are reported by our community.

2. Public reports conclude that drug cartels use social media like TikTok, META, X, Snapchat, and others to plan, organize, and communicate in real-time. These communications coincide directly with criminal activity. What are your companies doing to crack down on cartel coordination? Specifically, in the recruitment of children to commit crimes or assist in the sale/distribution of illicit drugs?

Nothing is more important than the safety of our community. We explicitly prohibit using Snapchat for illegal activity and use proactive detection tools to find this type of content and take it down. We closely collaborate with law enforcement to help prevent abuse on our service and encourage people to immediately report unlawful content, both through our in-app reporting tools and to law enforcement directly.

To help our community easily and quickly report any harmful or illegal content to us, we offer in-app reporting tools, as well as the ability to report via our Support Site or via Twitter. Reporting content is confidential, and allows us to preserve the reported content to investigate it.

As part of our ongoing work to help keep our community safe, we have an in-house Law Enforcement Operations team dedicated to reviewing and responding to law enforcement requests for data related to their investigations. Our global Trust and Safety teams also work around the clock to quickly investigate any reports and take appropriate action.

Snapchat has been working for years to remove drug dealers from our service. We combat this challenge in a number of ways:

- We block drug-related search terms and respond to those queries with educational content.
- We proactively detect and remove drug-related content including powders and pills.
- We preserve drug-related content to make it available for law enforcement and we make proactive referrals for prosecution.
- We've collaborated with other services and with NGOs to support some of the country's most significant public education campaigns so that people know one pill can kill.
- We launched signal sharing with Meta because we know criminals use multiple services and sharing signals can help us keep people safe.

We also support a legislative approach, including the Cooper Davis Act, which we have supported since its inception. Current law strictly limits when Snapchat and other online services can share information about attempts to abuse our services. The bipartisan Cooper Davis Act would establish a new legal framework for information sharing about the sale of fentanyl and other life-threatening drugs, expanding the ability of Snap and other providers to support law enforcement investigations and help bring perpetrators to justice. More can and must be done to combat this national crisis, and we look forward to continuing to work with the Senate to pass this important legislation.

3. What steps does your platform take to proactively remove, delist, and ban any posts, users, websites, and advertisements associated with the sale and distribution of fentanyl and other illicit drugs?

Please see response to question 2.

4. One area of growing concern is the sale and distribution of fake or counterfeit vaping devices online, particularly in connection with so-called Delta-8 THC. Counterfeit vapes, many coming from China, have unsafe and even potentially deadly chemicals. They have caused hospitalizations and death. What are your platforms doing to combat this problem?

Our Community Guidelines prohibit the illegal promotion of regulated goods or industries, including unauthorized promotion of gambling, tobacco or vape products, and alcohol. Our prohibition against illegal and regulated activities reflects our stalwart commitment to safety across Snapchat. Upholding these rules not only helps ensure our service is not misused for unlawful purposes, but also helps protect Snapchatters from serious harm. To help advance these aims, we partner extensively with safety stakeholders, NGOs, and law enforcement organizations to provide our community with educational resources and to generally promote public safety.

Additional guidance on prohibited illegal or regulated activities that violate our Community Guidelines is available here.

5. What are the main impediments your platform encounters in identifying all fentanyl and illicit drug advertisements posted to your platform(s) automatically? Please describe any circumstances in which you do not or cannot apply detection technologies against content transmitted on your platform(s).

Please see response to question 2.

6. How many posts, users, websites, and advertisements have you removed, delisted, and banned per year for the sale and distribution of fentanyl and other illicit drugs? How many per year? Have you seen an increase in illicit drugs being advertised to children on your platform(s)?

Last year, we removed approximately 2.2 million pieces of drug-related content and blocked approximately 705,000 associated accounts.

7. Are there any other roadblocks or impediments that you face in addressing fentanyl and illicit drug advertisements on your platform(s), and working with law enforcement on such matters? If yes, what are they? If not, how many cases have been transmitted to law enforcement and DEA?

We have long worked with DEA field agents who have been critical partners in our fight against fentanyl. We make proactive referrals to the DEA for prosecution and we work to share intelligence and signals to evolve our detection tools.

8. How do you work with organizations, advocates, and experts focused on drug prevention and addiction recovery to adapt your products and operations to keep up with the illicit drug crisis — including working with parents that have lost children due to lethal drugs bought online?

We are committed to educating both Snapchatters and the general public about the dangers of fentanyl. Our ongoing work to combat the nationwide fentanyl crisis includes education for the Snapchat community:

- When it comes to informing our community about the dangers of fentanyl, we meet young people where they are. Over the past two years, we have promoted in-app educational videos and news content warning about the dangers of counterfeit pills and directing them to credible resources from trusted experts. We also have in-app resources tailored to parents to give them greater visibility into their teens' online experiences.
- Our expert partners include the Centers for Disease Control and Prevention (CDC), the Substance Abuse and Mental Health Services Administration (SAMHSA), Community Anti-Drug Coalitions of America (CADCA), Truth Initiative and the SAFE Project.

Our commitment extends to educating the general public:

- We served as a Founding Partner of National Fentanyl Awareness Day to help raise awareness about the urgent crisis.
- As part of our ongoing efforts to raise public awareness, we teamed up with the Ad Council and other tech platforms on an unprecedented public awareness campaign to help people learn more about the dangers of fentanyl, and have partnered with Song for Charlie to reach young people where they are and educate them on about the hidden dangers of fake prescription pills laced with fentanyl. These resources have provided our community with information about the dangers of counterfeit pills and the importance of naloxone as a life-saving medication.

We also want to recognize the many families who have worked to raise awareness on these issues, pushed for change, and collaborated with lawmakers on important legislation like the Cooper Davis Act, which can help save lives.

9. What are the total number of meetings that your company has had with parents to address online safety concerns? Can you provide the total number of meetings over the last three years? Please separate this last question's answer by number per year.

Over the last three years, our team has met with more than twenty families to listen to their experiences and address online safety concerns. With some families, we meet more regularly – in some cases – monthly. Parents also serve on Snap's Safety Advisory Board (which meets three times a year) providing a valuable perspective on online safety. In addition, last year, our team participated in several national school safety conferences with an audience of parents, educators, and safety professionals. These included the National Association of School Resource Officers Annual Conference, the Safe and Sound Schools Annual Conference, three National Student

Safety & Security Conference and Workshops, as well numerous direct engagements with parents, school administrators, and school safety officers around the country.

10. In 2022, then National Center for Missing & Exploited Children (NCMEC) received over 32 million reports of Child Sexual Abuse Material (CSAM). Reports of online sex crimes to the CyberTipline are growing exponentially year by year. Out of those 32 million reports, how many did your platform submit to NCMEC?

Snap submitted approximately 551,000 CyberTips to NCMEC in 2022, and approximately 690,000 in 2023.

11. There is concern that this number is going to fall dramatically this year because of the adoption of end-to-end encryption, not because the problem is going away. How will your company track and address this issue moving forward?

We do not anticipate that our NCMEC reports will decrease this year. We believe encryption is important in helping people communicate with their friends and family privately and safely, but we do not plan to implement end-to-end encryption in a way that would prevent us from being able to detect known child sexual abuse material or constrain our ability to report it to authorities. We proactively scan for and detect known CSAM so we can immediately remove, investigate and report the content and violating accounts to authorities.

12. Has your platform seen an increase of suspected online child sexual exploitation- CSAM over the past few years? If so, what do you believe is the driving factor on why it's happening on your platform?

We are very much aware and focused on the growing rise in financially motivated "sextortion" — where criminals pose as young people and trick victims into sending compromising images. Because Snapchat is a platform that is attractive to teens and young adults, our platform is not immune to this trend that's affecting a range of online platforms. This conduct is illegal and abhorrent, and we take active steps to identify and prevent such exploitation as well as empowering users to identify and report suspicious contact.

We combat this criminal activity by:

- Making it difficult for strangers to search for, find, or contact teens.
- Warning teens if they receive a friend request from another person who isn't a mutual friend or phone contact.
- Keeping friend lists private and preventing teens from having public profiles — which helps prevent criminals from using Snapchat to target a teen's friends
- We proactively detect accounts attempting to engage in sextortion and disable them. When we remove sextortion-related content, we also retain it for an extended period of time in case law enforcement wants to follow up with a valid legal request.

- We've added an in-app reporting option specifically for sextortion ("They leaked/are threatening to leak my nudes.") to make it easier for our community to report such abuse to us. When we receive such a report, we take action quickly – usually within 30 minutes.
- We also educate teens about the dangers of this type of crime and urge them to use our in-app reporting tools.

This is very important because young people may feel afraid to report the problem to their parents or caregivers.

13. What are some new tools or strategies that your platform has implemented to identify CSAM? How closely does your platform work with NCMEC?

The sexual exploitation of any young person is horrific, illegal and against our policies.

We attack it in three key ways:

- We use cutting edge technology – PhotoDNA and CSAI Match – to proactively identify known CSAM photos and videos uploaded to Snapchat and report them to NCMEC.
- When we remove illegal and abusive content, we retain it for an extended period so that it is available if law enforcement want to follow up with a valid legal request.
- We take urgent action when we receive reports of CSAM, generally responding within 15 minutes.

We are actively exploring multiple privacy-protecting ways of utilizing technology like Google's Content Safety API to detect novel CSEAI. We will continue to iterate with our industry peers and expert organizations on ways to address these challenges.

14. What resources or help does your platform provide to victims of CSAM? Does your platform work with local victim groups and professionals?

As part of our overall effort to prioritize the mental health of Snapchatters, in 2020, we launched Here For You, a proactive in-app support system for Snapchatters who may be experiencing a mental health or emotional crisis by surfacing resources from expert organizations when Snapchatters search on a range of mental health-related topics, including eating disorders, anxiety, stress, depression, hate speech, bullying, suicidal thoughts, and more. When our Trust & Safety team recognizes a Snapchatter in distress, they will forward self-harm prevention and support resources, and notify emergency response personnel as appropriate. The resources that we share are available on our global list of safety resources, and these are publicly available to all Snapchatters.

15. What are the top technical hurdles your company faces in combating CSAM?

The creation and spread of AI-generated CSAM is quickly evolving and we welcome the collaboration of our industry partners, the government and law enforcement to combat this threat. AI-generated CSEAI, unfortunately, is already proving challenging for NCMEC and hotlines across the globe. Hotline analysts need to closely analyze photos for "tells" (e.g., extra digits, unnatural body positioning) as to whether the images may be wholly AI-generated or include partial depictions of real children. Either way, the process will take longer and delay referrals to law enforcement, as well as take away time from vital review cycles of real illegal imagery, depicting real children that need to be referred and actioned.

16. There seem to be competing views on how to regulate algorithms. Some suggest that more transparency is needed, while others want more privacy. Can you provide your perspective on whether more or less transparency is needed when it comes to algorithms?

Algorithms work differently on Snapchat. Unlike other services, we don't apply an algorithm to a feed of unvetted or unmoderated content and there is no rabbit hole of harmful content. Stories and Spotlight, the areas of our service where we apply algorithms to serve content, are closed services where content is moderated and we choose what content is approved for distribution.

Users can manage the type of content they may be served by adjusting their interest categories which are used to serve content we think they will prefer. We believe that our core architecture and design decisions limit the risk of algorithms that are applied to unmoderated feeds driven by engagement signals such as likes and comments. Across our service, we limit the risks of virality, which removes incentives for people to create content that appeals to people's worst instincts, and limits concerns associated with the spread of bad content such as disinformation, hate speech, self-harm content, or extremism.

17. Do you believe that large companies and platforms like yours can use algorithms to stifle innovation or small businesses?

We believe that algorithms can and should be designed to promote innovation and small businesses.

18. What do you believe is the role of the government in regulating algorithms? What, if any, unintended consequences would there be if Congress gets involved?

Congress plays a critical role in regulating technology and we believe the use of algorithms is a key piece of this oversight. The Kids Online Safety Act, which Snap supports, includes a provision on algorithms to ensure that the safety and well-being of young people is considered when providing algorithmically recommended content. Not all algorithms are intrinsically negative. One unintended consequence of mandating the ability to opt-out of algorithms could result in users seeing content that is not relevant or appropriate to them. We believe that Congress should focus on regulating harmful content and tread carefully when restricting the

algorithms that distribute such content to avoid unintended consequences.

19. Are you aware of your platform using surveillance advertisements to target children (anyone under the age of 18) with specific ads? If so, in your opinion, how can this be mitigated?

Advertisers have the ability to target advertising to different age groups as appropriate, and in accordance with the law, we do limit distribution of certain ads to certain groups to comply with applicable law (e.g. only targeting alcohol ads to 21+ in the US). For Snapchatters aged 13 to 17, we do not allow ads for things like prescription medication, gambling, alcohol, tobacco, sexualized content, weight loss products or dating apps. All ads on Snapchat must adhere to our Advertising Policies, Community Guidelines and Terms of Service. We have a moderation team that reviews ads for compliance, including ads that are reported by our community. Snapchat collects information on the types of content Snapchatters engage with and infers their interests in a limited number of non-sensitive interest categories. In our Settings menu, all Snapchatters can change these interest categories to better personalize their experience.

20. Beyond surveillance advertisements, are there any other algorithmic-based practices being implemented that are particularly detrimental to children? In your opinion, how can this be mitigated?

No.

21. Are you aware of any surveillance advertisements or algorithms that are used to target children, specifically to promote drugs and the sale of narcotics?

No.

Sen. Sheldon Whitehouse (D-RI)

1. What exemptions from the protections of Section 230 would your company be willing to accept?

Section 230 is a cornerstone of internet safety because it creates a legal mechanism that allows Internet companies to remove harmful content. If Section 230 didn't exist, internet companies wouldn't be able to effectively moderate their platforms to keep people safe. We wish more companies used Section 230 to remove harmful content. We believe that Section 230 could be amended to require, rather than simply allow, companies to remove specific types of harmful content that is not covered by the 1st Amendment, such as CSAM.

2. Is it your belief that your company should enjoy absolute immunity under Section 230 from suits like Doe v. Twitter, No. 21-CV-00485-JCS, 2023 WL 8568911 (N.D. Cal. Dec. 11, 2023), no matter the extent of your company's failure to remove reported child sexual abuse material from the platform or to stop its distribution?

Snap submits hundreds of thousands of reports of CSAM to NCMEC each year, and we take these legal obligations seriously. Snap would support legislation to ensure that platforms, upon receiving notice of CSAM on their platform, are obligated to remove that material within a reasonably prompt timeframe.

Sen. Cory Booker (D-NJ)

1. Trust and safety teams are a vital component in combating the spread of CSAM, hate speech, violence, and other violative content on tech platforms. Despite this, tech companies have time and time again disinvested from their trust and safety team, especially during changes in leadership.

> **a**. How has the size of your trust and safety team changed over the past five years? Please provide numbers for each of the past five years.

Our Safety & Moderation team has grown approximately 190% from 2019 to present.

> **b**. Do your trust and safety teams make submissions to the National Center for Missing and Exploited Children's CyberTipline, or is that a separate unit?

Yes, our Trust & Safety team makes submissions to NCMEC.

> **c**. If it is a separate unit, how many members are on the team and how have those numbers changed over the past five years. Please provide numbers for each of the past five years.

It is not a separate unit.

2. The National Center for Missing & Exploited Children's CyberTipline plays an integral role in combating child sexual exploitation. The tipline helps law enforcement investigate potential cases and allows prosecutors to bring justice to victims. While federal law requires your company to report to the CyberTipline any apparent violations of federal laws prohibiting child sexual abuse material of which you are aware, there are many gaps.

> **a**. Is there a standard format your reports to the CyberTipline follow? If so, what is that format?

Yes, we report to NCMEC through the CyberTipline API using the fields NCMEC provides for reporting.

> **b**. Does your company proactively report planned or imminent offenses?

Yes.

> **c**. Does your company proactively report potential offenses involving coercion or enticement of children?

Yes.

d. Does your company proactively report apparent child sex trafficking?

Yes.

Sen. Laphonza Butler (D-CA)

1. Family and parental control tools: I was glad to hear that you have spent time talking with parents and what their families need from your products. I was also glad to hear your companies have a Family Center, or other similar tools, to give parents more insight and control over how their children are using your platforms and apps.

a. How do you advertise this feature to parents?

We are constantly working to educate Snapchatters and parents about the accessibility of tools such as Family Center. We promote Family Center at large online safety conferences and events, as well as smaller gatherings and meetings. We engage with nonprofits and NGOs to help make parent groups and related organizations aware of Family Center, including with influential groups such as the National Parent Teacher Association (PTA) and the American Federation of Teachers. We also invest in search engine optimization advertising, allocating budget to Google search ads so that Family Center information is ranked at the top for parents looking for information on parental tools. We are also always looking for ways to make Family Center more easily accessible within Snapchat, to ensure parents have these tools at their fingertips.

b. Can you share data on how many Family Center/parental tools users there are in proportion to total minors on your platforms and products?

Approximately 200,000 parents use our Family Center suite of tools, and approximately 400,000 teen accounts have been linked to a parent's account through Family Center representing approximately 0.7% of our 60M global DAU aged 13-17.

Sen. Chris Coons (D-DE)

1. During the hearing, I asked the five witnesses whether the platform they represented publicly discloses "an estimate of the total amount of content—not a percentage of the overall…but the total amount of content on your platform—that violates" the platform's "policies prohibiting content about suicide or self-harm." I also asked if each platform "report[s] the total number of views that self-harm or suicide-promoting content that violates that policy gets on [each] platform." In response to these questions, you testified while under oath "Yes, Senator. We do disclose." After reviewing Snap's most recent transparency report from December 13, 2023, it appears that your testimony was misleading. First, Snap's transparency report does not disclose an estimate of the total amount of content on Snap's platform that violates the company's suicide and self-harm policy. Second, the transparency report does not disclose the estimated number of views that this violating content on the platform receives. These exclusions stand in direct contrast to your sworn testimony on January 31.

> **a**. Please provide the specific citation to where Snap publicly discloses an estimate of the total amount of content on the platform that violates Snap's suicide and self-harm policy.

We disclose the amount of content that violates the company's suicide and self-harm policy, and we disclose the violative view and viewer rate of such content which is an aggregate measurement of the views and viewers that this content received.

As part of our most recent Transparency Report, covering July 1, 2023 - December 31, 2023, we reported that 24,621 pieces of content and 22,637 unique accounts were enforced for suicide and self-harm, out of a total of 6,216,118 pieces of content and 3,687,082 that were enforced. We also note that this was 0.4% of total content enforced during this time-frame. The Transparency Report is publicly available at the following URL: https://values.snap.com/privacy/transparency.

In addition, our California transparency page links to our California Terms of Service Report, which was published at the time of the hearing. This report, in turn, includes a range of safety-related metrics, including providing the Violative View Rate (VVR) for individual categories of harms, as well as the Violative Viewer Rate which discloses how many viewers saw the content as a percentage of all active users over a period of time, both for the U.S. and globally. The report specifically indicates that suicide and self-harm content enforced by Snap's content moderators in Q3'23 had a violative view rate of approximately 0.000007% (and even less so for automated enforcements). Going forward, we plan to integrate this data into our main, semi-annual Transparency Report for convenience and clarity.

> **b**. Please provide the specific citation to where Snap publicly discloses an estimate of the total number of views of content that violates Snap's suicide and self-harm policy.

Please see response to question 1.a.

c. If Snap does not disclose these metrics, why not?

N/A

d. Does Snap measure these metrics? If not, why not?

Yes, we measure these metrics, as described in response to question 1.a.

2. Snap has previously reported how much content it removes under the platform's suicide and self-harm policy.

a. For content that has been removed, does Snap measure how many views that content received prior to being removed? If not, why not?

Yes, we measure this. Please see response to question 1.a for more information.

b. For content that has been removed, does Snap disclose how many views that content received prior to being removed? If so, please provide a specific citation to where Snap discloses that information. If not, why not?

We do not currently disclose this metric for each piece of content but rather disclose the violative view rate generally, as described in response to question 1.a. This statistic provides a more intelligible overview of the data, whereas if we were to disclose how many views all content received before they were removed, we'd have thousands of view metrics for each of the thousands of individual pieces of suicide and self-harm content that was enforced. If we extrapolate further, to all of our harms, that would be hundreds of thousands of view metrics, which would not provide meaningful clarity to the actions taken against our content, unlike VVR which provides a more meaningful point of comparison.

c. Please provide an estimate of the number of views content that was removed under this policy received in January 2024.

Snap estimates the VVR for self-harm and suicide at approximately 0.00002% for suicide and self-harm content in the U.S during January 2024.

d. For content that has been removed, does Snap measure demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If not, why not?

Snap measures violative views in the aggregate, via our Violative Viewer Rate, as described in response to question 1.a. The Violative Viewer Rate is the percentage of unique viewers who saw violating content, as a proportion of unique users active throughout the reporting period — but this metric does not currently differentiate between demographics, such as whether the unique user was a minor or an adult, in part because we are committed to reducing violative views of potentially harmful content regardless of users' demographics.

> **e**. For content that has been removed, does Snap disclose demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If so, please provide a specific citation to where Snap discloses that information. If not, why not?

No, we do not, for similar reasons to that described in response to question 2.b.

> **f**. Does Snap measure the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If not, why not?

Yes, we do, as described in response to question 1.a.

> **g**. Does Snap disclose the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If so, please provide a specific citation to where Snap discloses that information. If not, why not?

While we do not disclose specifically the number of users that have viewed a specific piece of content multiple times or, generally, viewed violative content multiple times for reasons discussed in response to question 2.b, we disclose statistics relating to violative views and viewers, as discussed in response to question 1.a.

3. Snap utilizes an algorithm to recommend or amplify content to users.

> **a**. For content that has been removed, does Snap measure whether and the extent to which the removed content was recommended or amplified by Snap? If not, why not?

This question assumes that Snapchat algorithmically recommends content that may violate our Community Guidelines. However, on Snapchat's public content surfaces—including Spotlight and Snapchat Stories (also known as Discover)—we moderate content before it can be recommended to a large audience. We also use advanced technology (such as signal-based detection, machine learning, and keyword lists) to proactively identify and remove certain types of harmful or illegal content upon submission (including CSAM, illicit-drug, and sextortion-related content). If content on these surfaces—or elsewhere in Snapchat—violates our Community Guidelines, we at a minimum remove it and it is not eligible for recommendation. If content violating our Community guidelines does slip through our moderation process, we

remove it if it is later reported or discovered, but do not track the extent to which the removed content was recommended or amplified.

> **b**. For content that has been removed, does Snap disclose whether and the extent to which the removed content was recommended or amplified by Snap? If so, please provide a specific citation to where Snap discloses that information. If not, why not?

As discussed in response to question 3.a, algorithmically-recommended content is moderated. In Discover, Publishers and Snap Stars are verified by our internal partnerships team to maintain content quality standards. In Spotlight, content is moderated according to multiple policies, including our Content Guidelines for Recommendation Eligibility, Community Guidelines, and Terms of Service. On both surfaces, content that is reported is reviewed by our moderation team for compliance against our Content Guidelines for Recommendation Eligibility. These intentional design choices minimize the risk of violative content from going viral.

> **c**. For content that has been removed, does Snap measure how many views the removed content received after having been recommended or amplified? If not, why not?

Yes, Snap stores the number of views removed content received before being removed.

> **d**. For content that has been removed, does Snap disclose the number of views the removed content received after having been amplified or recommended? If so, please provide a specific citation to where Snap discloses that information. If not, why not?

No. Given our robust moderation policies, there is a limited amount of violative content that goes viral.

4. Does Snap support creating industry-wide transparency requirements to disclose basic safety information, like those included in the Platform Accountability and Transparency Act?

It is our understanding that the Platform Accountability and Transparency Act would allow researchers to access personal data including potentially sensitive personal information. While we are supportive of industry-wide transparency requirements for basic safety information, we take seriously our obligations to protect user data and would not support research access for personal data and sensitive personal information without explicit consent.

Sen. Ted Cruz (R-TX)

Directions

Please provide a wholly contained answer to each question. A question's answer should not cross-reference answers provided in other questions.

If a question asks for a yes or no answer, please provide a yes or no answer first and then provide subsequent explanation.

If the answer to a yes or no question is sometimes yes and sometimes no, please state such first and then describe the circumstances giving rise to each answer.

If a question asks for a choice between two options, please begin by stating which option applies, or both, or neither, followed by any subsequent explanation.

If you disagree with the premise of a question, please answer the question as-written and then articulate both the premise about which you disagree and the basis for that disagreement.

If you lack a basis for knowing the answer to a question, please first describe what efforts you have taken to ascertain an answer to the question and then provide your tentative answer as a consequence of its reasonable investigation.

If even a tentative answer is impossible at this time, please state why such an answer is impossible and what efforts you intend to take to provide an answer in the future.

Please further give an estimate as to when Senator Cruz will receive that answer. To the extent that an answer depends on an ambiguity in the question asked, please state the ambiguity you perceive in the question, and provide multiple answers which articulate each possible reasonable interpretation of the question in light of the ambiguity.

1. In the last two years, has an employee or commissioner of the Federal Trade Commission (FTC) requested to evaluate or evaluated your data used for training Large Language Models or algorithms or the sources of such data for bias, discrimination, or misinformation?

No.

2. In the last two years, has an employee or commissioner of the FTC sought details regarding your company's measures related to filtering or blocking inputs and outputs of a Large Language Model or algorithms?

Yes, the FTC has sought details regarding how Snap trains its chatbot, My AI, to control or modify outputs generated by the chatbot.

> **a**. If yes, has the FTC attempted to coerce or otherwise request you to implement input/output filtering in order to allegedly comply with federal law?

No.

3. In the last two years, has an employee or commissioner of the Federal Trade Commission sought to evaluate your company's use of measures, including "prebunking" or "debunking", designed to counteract so called "online misinformation?"

No.

4. In June 2022, the FTC released a report titled "Combatting Online Harms Through Innovation." In this report, the FTC discussed how the deployment of AI tools intended to detect or otherwise address harmful online content is accelerating but may never be appropriate as an alternative to human judgment.

> **a**. In the context of protecting children from online harms to what extent does your company rely on automated tools to detect online harm vs. human review? Please be specific.

To help prevent the spread of harmful content on Snapchat, we use a combination of human and automated tools to enforce our Community Guidelines. For example:

- We use automated mechanisms such as machine learning, heuristics-based rule engines and hash matching technologies to proactively detect violative content and accounts, and to take enforcement action. When these mechanisms have high enough confidence they are used to make automated decisions and, when unsure, content and accounts detected by these models are augmented with human review labels to make enforcement decisions.
- We similarly use a combination of automation and human review to approve the content that is recommended on our service.
- We use human review decisions and labels as the source of truth to evaluate the efficacy of our automated mechanisms, perform quality assurance checks and as critical inputs to augment automation.
- We use automated abusive language matching mechanisms to block search results for a wide range of terms related to online harms. In some instances, such as with drug-related terms, we instead redirect Snapchatters to resources from experts about the dangers of fentanyl.

- We use human review processes to fact-check all political and advocacy ads. All political ads, including election-related ads and issue advocacy ads, must include a transparent "paid for" message that discloses the sponsoring organization, and we provide access to information about all ads that pass our review in our Political Ads Library.

b. What benefits can AI provide to helping detect and/or stop harmful content to children online?

Automated tools, including AI and Machine Learning systems, represent an important part of our overall approach to moderation. For example, we use machine learning to specifically train models to proactively detect harmful content or to identify suspicious account behavior — for example relating to illicit drug content or distribution. We do this so we can immediately remove such offending content from Snapchat. The ability to improve our proactive scanning capabilities with artificial intelligence is critical to our safety efforts.

c. What does a human reviewer provide that an AI or automated tool cannot? Will we always need some measure of human review in assessing online harms to children?

We believe a combination of human review and machine learning or AI technologies provides the most comprehensive approach to moderation. We continue to invest in new machine learning capabilities to improve our proactive detection of potentially harmful content. At the same time, human review is critical to accurately assess and take action on reported or proactively detected content at least until automated machine learning performs on par with human review of the particular type of content. In addition, human review is necessary to label and perform quality assurance on tasks generated by machine learning models. We have actually grown the size of our human review teams significantly in recent years.

d. The FTC has sent mixed signals in its enforcement of COPPA. While the Commission emphasizes not over relying on use of automated tools or AI, they have nonetheless found liability for using human review as alternative signaling overreliance on automated tools. What improvements, if any, should Congress make to clarify the legal tension between use of automated detection tools vs. human review?

We believe clearer guidance and consistent enforcement from the FTC on human review and use of automated tools or AI is needed. Additional guidance from Congress providing legislative direction to the FTC to not penalize companies who use human review methods in combination with automated tools would be valuable and ensure that companies who make good faith efforts to keep their service safe are not unfairly punished.

5. In 2021, Congress directed the FTC to research and report on how AI can be used positively to detect and combat fraudulent or deceptive content online. Rather than viewing AI as a potential

solution to our online woes, the FTC instead issued a report that read more like an indictment of the technology.

> **a**. Please explain whether, in your view, AI can be used to positively detect and combat fraudulent or deceptive content, including the recent use of deep fakes or other scams to harm consumers.

We have observed that AI and other Machine Learning technologies hold enormous promise in supporting comprehensive efforts to identify and combat the spread of harmful content. While automated tools are just one element of a multifaceted, multilayered safety strategy at Snap, advances in AI technologies (and adjacent developments in the realm of content provenance and authenticity) are unlocking new capabilities in support of critical harm mitigation efforts.

> **b**. Has the FTC ever consulted with your company to learn how your company deploys AI to better detect and combat fraudulent or deceptive content? Has the DOJ? How about the Federal Elections Commission?

Yes. The FTC has issued an order under 6(b) of the FTC Act to eight social media and video streaming services, including Snap, seeking information on how the services scrutinize and restrict paid commercial advertising that is deceptive or exposes consumers to fraudulent healthcare products, financial scams, counterfeit and fake goods, or other fraud. As part of this inquiry, the FTC asked for information about how the services use algorithmic, machine learning, or other automated systems to detect potentially misleading, deceptive, or fraudulent advertisements submitted for publication on Snapchat.

> **c**. How can Congress empower agencies to use AI positively for the protection of American consumers from fraudulent or deceptive content?

We anticipate that different federal agencies will each have distinct uses for AI that can positively advance their mission and help protect consumers; our team would be eager to work with your office to identify high-value opportunities to support these uses.

6. Please provide a description of your company's policy regarding the sale or transfer of the data of American users collected on your platform to a third party, including data brokers.

We do not sell user data. We may share data of our users for legitimate business reasons, for example, to cloud storage providers to store user data, and other service providers to enable the services we offer to our users. We detail our practices in our Privacy Policy.

7. Has your company ever sold the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the categories of data sold.

No.

8. Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the circumstances underlying the basis for such transfer.

No, we do not voluntarily transfer data of American users to the government of a foreign country except in limited circumstances where the information relates to an emergency involving danger of death or serious physical injury, which requires the disclosure, as permitted by the U.S. Stored Communications Act.

9. Has your company ever sold the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and the categories of data sold.

No.

10. Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and categories of data transferred.

No, we do not voluntarily transfer data of American users to U.S. government agencies except in limited circumstances where the information relates to illegal conduct or an emergency involving danger of death or serious physical injury, which requires the disclosure, as permitted by the U.S. Stored Communications Act.

11. Does your company have a policy to restrict third party use and/or transfer of data collected from users on your platform? Please be specific, including how you enforce such restrictions and whether such restrictions prohibit the sale or transfer of such data to a government agency, including a foreign government agency.

Yes, we restrict the sharing of data with third parties, as detailed in our Privacy Policy. We enforce the Privacy Policy and data sharing restrictions in our privacy-by-design and legal review process, which includes a prohibition on the sale of user data.

12. Between July 4, 2023 and July 14, 2023, was your company contacted by any employee of or contractor for any of the following agencies? Please answer "yes" or "no" for each agency and, if "yes," provide the date(s) of contact and the name(s) of the agency employees or contractors that contacted your company.

a. U.S. Department of Health and Human Services (HHS)

No.

b. National Institute of Allergy and Infectious Diseases (NIAID)

No.

c. Centers for Disease Control and Prevention (CDC)

No.

d. U.S. Food and Drug Administration (FDA)

No.

e. The National Institutes of Health (NIH)

No.

f. U.S. Department of Homeland Security (DHS)

Yes. We regularly respond to valid legal requests from employees of the DHS' investigative agency, Homeland Security Investigations.

g. DHS Cybersecurity and Infrastructure Security Agency (CISA)

No.

h. U.S. Census Bureau

No.

i. Federal Bureau of Investigation (FBI)

Yes. We regularly respond to valid legal requests from employees of the FBI.

j. U.S. Department of Justice (DOJ)

Yes. We regularly respond to valid legal requests from employees of the DOJ.

k. The White House Executive Office of the President (EOP)

No.

1. U.S. Department of State

No.

13. Is it your company's policy to prevent children under 13 from using your social media app(s) or creating an account?

Yes. When registering a Snapchat account, users are required to accept Snap's Terms of Service, acknowledge the Privacy Policy, and provide their date of birth. Users need to be at least 13 to create a Snapchat account, and the registration process is blocked if a user inputs a birthdate that indicates they are under the age of 13.

If we become aware that a Snapchat user is under the age of 13, we will terminate that user's account and delete the user's data. We also implement a safeguard that prevents younger Snapchat users with existing accounts from updating their birthday to an age of 18 or above. Specifically, if a Snapchat user between the age of 13 to 17 attempts to update their year of birth to reach an age over 18, we will prevent the change.

14. In your view, would it be appropriate for school-aged children to spend time on or access your company's social media app(s) during class?

Yes, if it is to communicate with their parents or caregivers regarding an urgent matter.

15. As a parent, would you be concerned if your child were able to access your company's social media app(s) during class via a school network or device?

No.

16. In your view, should elementary and secondary schools block students' access to your company's social media app(s) on school networks and devices?

No. I am concerned this would unreasonably restrict teen students' ability to communicate with their family and friends.

17. Do you think that school buses equipped with Wi-Fi should allow children to access your company's social media app(s) via a school bus Wi-Fi network during their rides to and from school?

Yes. Snapchat is an important communications service for parents and their teens, and using Snapchat before or after school or during a bus ride does not interfere with the classroom environment.

18. As a parent, do you think it is important to supervise your children's internet access?

Yes.

19. As a parent, would you be concerned if your child's school allowed your child to access the internet on an unsupervised basis, such as on your child's bus ride to and from school via the school bus Wi-Fi?

No. Our family utilizes parental controls that exist on the device level that manages the kind of access and content our child is permitted to view whether using Wi-Fi or cellular data.

20. Do you think Congress should require schools, as a condition of receiving broadband subsidies through the Federal Communications Commission's E- Rate program (which funds broadband for elementary and secondary schools), to block students' access to your company's social media app(s) from school-run networks?

No. We believe this decision is better left to parents and schools.

21. Do you support the bipartisan Eyes on the Board Act of 2023, S. 3074?

Snap has not taken a formal position on this piece of legislation.

22. Have you, your company, or any foundation associated with you or your company, donated or contributed funding, equipment, or services to any of the following organizations in the last ten years (CY 2013 to CY 2023)?

a. Education and Libraries Networks Coalition (EdLiNC)
b. Open Technology Institute
c. Consortium for School Networking (COSN)
d. Funds For Learning
e. State Educational Technology Directors Association (SETDA)
f. Schools, Health, and Libraries Broadband Coalition (SHLB)
g. State E-Rate Coordinators' Alliance (SECA)
h. EducationSuperHighway
i. All4Ed
j. Public Knowledge
k. Fight for the Future
l. Free Press

m. Electronic Frontier Foundation
n. Benton Foundation or Benton Institute for Broadband & Society
o. Electronic Privacy Information Center

No, we have no record of any donations made to these organizations.

23. For each such donation or contribution described in the prior question, please detail (1) the type of donation or contribution, such as financial donation, goods or equipment, services, etc.; (2) who made the donation or contribution; (3) the recipient organization; (4) the year the donation or contribution was made; and (5) the total value of that donation or contribution.

N/A

Senator Dick Durbin
Chair, Senate Judiciary Committee
Written Questions for Linda Yaccarino
Chief Executive Officer, X Corp.
February 7, 2024

1. For each year from 2019 to 2023, please provide the following:

   a. the total number of users on your platform;

   2019 - Average DAU (Daily Active Users) was approximately 152 million
   2020 - Average DAU was approximately 192 million
   2021 - Average DAU was approximately 217 million
   2022 - Average DAU was approximately 238 million
   2023 - Average DAU was approximately 245 million

   b. the total number of users under the age of 18 on your platform;

   2019 - 2022 - We will follow up with this figure.
   2023 - As of January 2024, Average DAU of users 13-17 in US is approximately 650,000

   c. the estimated number of users under the age of 13 on your platform;

   X does not allow individuals under the age of 13 to open an account on the platform. We do not have estimates for the number of users under the age of 13.

   d. your company's annual revenue;

   Revenue for the years 2019, 2020, 2021 and first two quarters of 2022 were disclosed to the SEC as part of Twitter's public company filings.

   2019 revenue was $3.46 billion, with an operating income of $366 million, and net income of $1.47 billion.

   2020 revenue was $3.72 billion, with an operating income of $27 million, and net loss of $1.12 billion.

   2021 revenue was $5.08 billion, with a 2021 operating loss of $493 million and net loss of $221 million.

   2022 Q1 revenue was $1.2 billion, an operating loss of $128 million, and a net income of $513 million which includes a pre-tax gain of $970 million from the sale of MoPub for $1.05 billion and income taxes related to the gain of $331 million.

   2022 Q2 revenue was $1.18 billion, with an operating loss of $344 million, and a net

loss of $270 million.

Beginning in Q3 2022, X Corp., a private company, took over operation of the platform. As a privately-held company, X does not maintain or release public financial statements.

e. your company's annual budget for trust and safety;

X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.

f. your company's annual budget to address online child sexual exploitation;

X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.

g. the total number of individuals at your company working to address trust and safety broken down between employees and contractors;

We have approximately 2,300 people who work on Trust & Safety and content moderation.

h. the total number of individuals at your company working to address online child sexual exploitation broken down between employees and contractors.

X has a combination of program managers, policy specialists, operations specialists, engineers, legal professionals, government affairs professionals, and other functions that work on issues related to child safety.

2. How did your company determine that 13 was the appropriate age for a child to begin using your platform?

Our age limit is in alignment with the Children's Online Privacy and Protection Act.

3. What legal obligation does your company have in the United States to ensure that your platform or features of your platform are safe for children before they are launched?

The Children's Online Privacy and Protection Act, as well as guidance from the Federal Trade Commission, provide a framework for ensuring safety and privacy of minors online.

4. For users under the age of 18,

a. what are the default privacy settings for their accounts?

Accounts belonging to known minors will be defaulted to a "protected" setting. This

means that known minors will receive a request when new people want to follow them (which they can approve or deny), that their posts will only be visible to their followers, and that their posts will only be searchable by them and their followers (i.e. they will not appear in public searches). Under this setting, accounts belonging to known minors will be restricted to receiving DMs from accounts they follow by default. We also utilize an age lock. Once a new user enters a date of birth that makes them under the age of 18, they will be stopped from re-entering a new date of birth for that account.

We also take steps to limit exposure to sensitive content. Known minors or viewers who do not include a birth date on their profile are restricted from viewing specific forms of sensitive media such as adult content. X obscures sensitive media behind notices and interstitials. This includes our product age restrictions that restricts known minors from viewing adult content.

In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.

More information on our protected account settings can be found in our Help Center: https://help.twitter.com/en/safety-and-security/public-and-protected-posts.

b. what limitations are placed by default on content these users can access, content that will be directed toward them, and individuals they can communicate with?

See answer above.

c. can they change their default settings without the awareness of their parent or guardian, or without the consent of their parent or guardian?

Yes.

d. In 2023, how many changed their default settings?

In response to your inquiry, our teams are investigating this question and will be happy to follow up with you.

5. If default settings are different for users aged 16 and 17 than they are for users under the age of 16, please explain why these groups of users are treated differently, how the decision to treat these groups of users differently was made, and whether any company personnel voiced objections to or raised concerns about the differing treatment of these groups of users.

The default settings are not different for users between the ages of 13-17.

6. Please describe what parental controls, if any, are available on your platform. What studies, research, summaries, or data does your company have reflecting the efficacy of its parental controls and child safety measures? Please provide these studies, research, summaries, or data.

   X does not currently have parental controls, however, we will be engaging with parents and parents groups to solicit feedback on what tools could be helpful to develop. While there are not a great amount of kids and young teens on X, there are a lot of parents that we can learn from and involve in designing new products and solutions.

7. **Concerning international law,**

   a. what steps have your company and its subsidiaries taken to comply with the European Union's *Digital Services Act*?

      A number of steps have been taken with regard to the Digital Services Act, including risk assessments, compliance process, introduction of a specific reporting process, the publication of an EU transparency report, and processes relating to academic data access. We continue to work closely with the European Commission to detail our compliance with the DSA.

   b. what steps have your company and its subsidiaries taken to comply with the United Kingdom's *Online Safety Act*?

      The Online Safety Act has not yet come into force, but we are engaging closely with the regulator, OFCOM, and are evaluating any potential policy and product changes required.

   c. what steps have your company and its subsidiaries taken to comply with Australia's *Online Safety Act*?

      X complies with the Online Safety Act and continues to work diligently to cooperate in good faith with the Australian eSafety Commissioner. Given ongoing litigation, we decline to comment further at this time.

   d. if those laws create a safer, healthier online experience for kids on your platforms, do you commit to implement these changes for users in the United States? If not, why not?

      In many areas, the work that we do to comply with these laws will have a global impact, for example our investments in tackling child sexual exploitation content. Improvements in our reporting flow and greater transparency relating to our content moderation work have been rolled out globally.

8. X claims to prioritize reducing child exploitation on its platform. In fact, in November 2022, Elon Must tweeted "Priority #1" in response to a tweet about the company addressing child sexual exploitation content.

Yet, within the past two years, X reduced its global trust and safety staff by 30 percent, including 80 percent of its staff engineers. The head of X's trust and safety team also resigned from the company in 2023.

While this was happening, *NBC News* reported that "at least dozens of accounts have continued to post hundreds of tweets in aggregate using terms, abbreviations and hashtags indicating the sale of... child sexual exploitation material." Researchers at Stanford University similarly found that X failed to prevent dozens of known images of child sexual abuse from being posted on its platform in recent months.

How is X prepared to fight child exploitation on its platform with an understaffed and under-resourced trust and safety team?

Since acquisition, we have invested in technology, training, and people to strengthen our approach to combating child sexual exploitation on X. In 2023, as a result of our investment in additional tools and technology to combat CSE, X suspended 12.4 million accounts for violating our CSE policies. This is up from 2.3 million accounts in 2022.

Not only are we detecting more bad actors faster, we are also building new defenses that proactively reduce the discoverability of posts that contain this type of content. One such measure that we have recently implemented has reduced the number of successful searches for known Child Sexual Abuse Material (CSAM) patterns by over 99% since December 2022.

We are investing in products and people to bolster our ability to detect and action more content and accounts, and are actively evaluating advanced technologies from third-party developers that can enhance our capabilities.

In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year. For the first time ever, we are evaluating all videos and GIFs posted on X for CSAM. Since launching this new approach in July 2023, we have matched over 70,000 pieces of media.

We are more vigilant and aggressive than ever in our enforcement. Our team regularly reviews and implements improvements to the measures we take to combat online child sexual exploitation to ensure their ongoing efficacy and performance. Our increased investment in this area throughout the year has yielded significant, measurable results.

Since April, we have increased training for content moderators on the tools and policies for NCMEC reporting. In turn, this has led to a 10x increase in the volume of manually-submitted NCMEC reports, from an average of 6,300 reports per month to an

average of 64,000 reports per month from June through November 2023. We are evaluating more sources of potential CSAM than we could before.

While reports focus on the cuts we made over a year ago, we are hiring across the company and we are excited about building a Trust & Safety Center of Excellence in Austin, Texas. We have a goal of hiring 100 Trust & Safety agents, moving more capacity in-house, relying less on contractors. This shift will increase efficacy, decrease turnover, develop more specialization across issue areas, improve quality assurance, and strengthen our enforcement. We want to work with Congress and all stakeholders to help develop our capacity here in the United States for this important work–we stand ready to partner on this workforce development challenge.

9. At the hearing, you repeatedly said that X is not a platform of choice for youth. However, you recently stated at a forum that Gen Z is X's fastest-growing demographic, with 200 million teenagers and young adults visiting the platform each month. You also recently noted that 70 percent of X's growth in the prior six months was driven by Gen Z users.

   What is your company doing to prevent the grooming of young people joining X?

   The growth of Gen Z users is largely in the older portion of the generation in their 20s, as our user base of minors between the ages of 13-17 has decreased over the last year.

   We are currently beta testing a new text-based machine-learning classifier to detect different types of child abuse discussion, which has yielded enforcement for grooming behavior. We look forward to sharing more as we continue to develop and implement this technology.

10. In an October meeting, Elon Musk highlighted a new feature on X that allows paying users to upload hours-long videos. You recently said that "the strides we are making so quickly in video" are "certainly getting everyone's attention."

    What steps has X taken to ensure its new video features are safe for the growing number of young people on your platform?

    People use X to show what is happening in the world, often sharing images and videos as part of the conversation. Sometimes, this media can depict sensitive topics, including graphic content, adult nudity, and sexual behavior. We recognize that some people may not want to be exposed to sensitive content, which is why we balance allowing people to share this type of media with helping people who want to avoid it to do so.

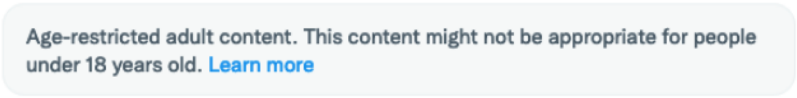
    For this reason, you can't include graphic content, adult nudity, or sexual behavior within areas that are highly visible on X, including in live video, your profile picture or header, List banners, or Community cover photos. If you share this content on X, you need to mark your media or your account as sensitive. Doing so places images and videos behind a content warning that needs to be acknowledged before your media can be viewed. Using this feature means that people who don't want to see sensitive media

can avoid it, or make an informed decision before they choose to view it. We also restrict graphic media, adult nudity, and sexual behavior for viewers who are under 18 or viewers who do not include a birth date on their profile. Beginning January 2024, you may begin to see new media content warnings on posts that X has designated as Graphic (containing violent or hateful imagery) or containing Adult media (adult nudity and sexual behavior). When these new content warnings are available for you to use, please be sure to continue marking your sensitive media accordingly.

Under this policy, there are also some types of sensitive media that we don't allow at all, because they have the potential to normalize violence and cause distress to those who view them.

We restrict viewers who are under 18, or who do not include a birth date on their profile, from viewing adult content.

Below is an example of the interstitial we use to restrict sensitive content:



11. An October 2023 report from Australia's eSafety Commissioner revealed that X does not employ language analysis technology to detect likely online grooming. This sets the company apart from other tech companies like Google, TikTok, and Twitch. In defense, X claimed that, although it continues to monitor the development of such technology, it is not of "sufficient capability or accuracy to be deployed by [X]."

    Could you elaborate on why language analysis technology is not capable or accurate for use on X when other platforms have demonstrated otherwise?

    X is currently beta testing a text-based machine learning classifier developed by Thorn to assist in the detection of sextortion, CSAM, child-access, and child sexual abuse discussion.

    Last year, following our submission to the Australian regulator, we developed and implemented our first CSE text model. The CSE model detects text-based posts that discuss CSE. We launched this model in February 2023 and it has been critical in addressing CSE spam. Since launching, we have restricted its impact to only a subset of languages and users due to several false positives, but it continues to be our most flexible means of CSE text detection with 65,000 total suspensions so far.

12. An impacted parent provided a statement to the Committee. She describes reporting CSAM depicting her son to X, only to be told that the images did not violate their policy. When she sued X, the lawsuit was dismissed because of Section 230.

When an individual reports child sexual abuse material on X, how do you resolve disputes as to whether the content violates X's policies? How long does that process take?

Anyone can report potential CSAM, whether they have an X account or not.

In the majority of cases, the consequence for violating our CSAM policy is immediate and permanent suspension from the platform. In addition, violators will be prohibited from creating any new accounts in the future.

Additionally, when we are made aware of content depicting or promoting child sexual exploitation, including links to third party sites where this content can be accessed, we immediately remove it without further notice and report to the National Center for Missing & Exploited Children (NCMEC) where appropriate.

In a limited number of situations, where we haven't identified any malicious or sexually exploitative intent, we will require the user to remove this content. We will also temporarily lock the user out of their account before being able to post again. Further similar violations lead to the account being permanently suspended.

We review 100% of reports for child sexual exploitation and take immediate action on confirmed hash matches. For media-based violations, since it is visual in nature, it is usually easy to identify without needing language expertise. For text-based violations we have dedicated employees with different language expertise who are reviewing both written and media-based content. We have dedicated training resources for agents who review CSAM.

We have quality assurance processes and dedicated specialists that help us identify gaps in policies and enforcement. Our content moderators provide moderation services 24 hours a day, 7 days a week. We have teams spread around the world specifically trained in this highly sensitive and complex topic so that we can provide the best possible level of coverage in the languages we serve on X. For safety and security reasons, the locations of these teams are not disclosed. We have a dedicated tool that not only takes action on content, but also has the ability to dispatch communication to NCMEC. This allows us to report CSAM to NCMEC as fast as possible.

13. You testified that X has a zero-tolerance policy for child sexual exploitation and that users who violate this policy face immediate and permanent suspension. Last summer, X suspended the account of an individual who tweeted an image of a toddler being tortured. When the account was suspended, Elon Must tweeted that the account, "was suspended for posting child exploitation pictures." Four days later, X reinstated the account. The individual who produced that child sexual abuse material was later sentenced to 129 years in prison for sexually abusing children as young as 18 months old. The image has now drawn more than 3 million views and 8,000 retweets.

    Please explain how this reflects X's zero-tolerance policy.

Our public policy sets out how in the rare circumstances where content is shared by an account from the perspective of outrage or to raise awareness, we will remove the content and give the account holder a final warning. Any further violations will result in permanent suspension.

We followed our publicly stated policy, and the user was given a final warning. The user was also reported to NCMEC, which triggered the initial suspension.

Senator Lindsey O. Graham
Questions for the Record
Ms. Linda Yaccarino, CEO, X Corp.
"Big Tech and the Online Child Sexual Exploitation Crisis"
January 31, 2024

1. Do you support S. 1207, the bipartisan EARN IT Act? Why or why not?

We support the concepts within EARN IT that encourage the development of best practices to combat the distribution of CSAM. We welcome more mechanisms for industry, law enforcement, and government to share these best practices and increase collaboration. We do not believe that Section 230 protections should be conditioned on the implementation of these best practices. We also have concerns about creating an unelected commission that can become politically charged with the changing of administrations.

2. What measures are you taking to prevent and address sextortion, including financial sextortion, on your companies' platforms?
   a. What methods are in place to detect and disrupt this type of abuse in real time?

We use a mixture of proactive detection, user reporting, and human capacity to enforce our private information and media policy.

You may not publish or post other people's private information without their express authorization and permission. We also prohibit threatening to expose private information or incentivizing others to do so.

In addition, you may not share private media, such as images or videos of private individuals, without their consent. However, we recognise that there are instances where users may share images or videos of private individuals, who are not public figures, as part of a newsworthy event or to further public discourse on issues or events of public interest. In such cases, we may allow the media to remain on the platform.

Sharing someone's private information online without their permission, sometimes called doxxing, is a breach of their privacy and of the X Rules. Sharing private information can pose serious safety and security risks for those affected and can lead to physical, emotional, and financial hardship.

When reviewing reports under this policy, we consider a number of things, including:

- ***What type of information is being shared***
  - **We take this into consideration because certain types of private or live information carry higher risks than others, if they're shared without permission. Our primary aim is to protect**

individuals from potential physical harm as a result of their information being shared, so we consider information such as physical location and phone numbers to be a higher risk than other types of information. We define "live" as real-time and/or same-day information where there is potential that the individual could still be at the named location.

- ***Who is sharing the information***
  - **We also consider who is sharing the reported information and whether or not they have the consent of the person it belongs to. We do this because we know that there are times when people may want some forms of their personal information to be shared publicly. For example, sharing a personal phone number or email for professional networking or to coordinate social events or publicly sharing someone's home addresses or live locations to seek help after a natural disaster.**
- ***Whether the information available elsewhere online***
  - **If the reported information was shared somewhere else before it was shared on X, e.g., someone sharing their personal phone number on their own publicly accessible website, we may not treat this information as private, as the owner has made it publicly available. Note: we may take action against home addresses being shared, even if they are publicly available, due to the potential for physical harm.**
- ***Why the information is being shared***
  - **We also factor in the intent of the person sharing the information. For example, if we believe that someone is sharing information with an abusive intent, or to harass or encourage others to harass another person, we will take action. On the other hand, if someone is sharing information in an effort to help someone involved in a crisis situation like in the aftermath of a violent event, we may not take action. Note: regardless of intent, if the information is not shared during a crisis situation to assist with humanitarian efforts or in relation to public engagement events, we will remove any posts or accounts that share someone's live location.**
- ***Sharing private media***
  - **Posting images is an important part of our users' experience on X. Where individuals have a reasonable expectation of privacy in an individual piece of media, we believe they should be able to determine whether or not it is shared. Sharing such media could potentially violate users' privacy and may lead to emotional or physical harm. When we are notified by individuals depicted, or their authorized representative, that they did not consent to having media shared, we will remove the media. This policy is not applicable to public figures.**

What is in violation of this policy?

Under this policy, you can't share the following types of private information, without the permission of the person who it belongs to:

- **home address or physical location information, including street addresses, GPS coordinates or other identifying information related to locations that are considered private;**
- **live location information, including information shared on X directly or links to 3rd-party URL(s) of travel routes, actual physical location, or other identifying information that would reveal a person's location, regardless if this information is publicly available;**
- **identity documents, including government-issued IDs and social security or other national identity numbers – note: we may make limited exceptions in regions where this information is not considered to be private;**
- **contact information, including non-public personal phone numbers or email addresses;**
- **financial account information, including bank account and credit card details;**
- **other private information, including biometric data or medical records;**
- **media of private individuals without the permission of the person(s) depicted; and**
- **media depicting prisoners of war posted by government or state-affiliated media accounts on or after April 5, 2022.**

The following behaviors are also not permitted:

- **threatening to publicly expose someone's private information;**
- **sharing information that would enable individuals to hack or gain access to someone's private information without their consent,e.g., sharing sign-in credentials for online banking services;**
- **asking for or offering a bounty or financial reward in exchange for posting someone's private information;**
- **asking for a bounty or financial reward in exchange for not posting someone's private information, sometimes referred to as blackmail.**

3. Please provide the committee statistics on how long it takes your company to respond to various types of legal process from law enforcement?

X endeavors to respond to legal process received from law enforcement and appropriate government entities in a prompt manner. Specifically, X endeavors to respond prior to the enumerated production date required by law in the particular jurisdiction or outlined in the individual legal process.

4. Do you notify your users when law enforcement serves subpoenas/summons for subscriber information and specifically requests not to notify the subscriber/user?

Upon receipt of a law enforcement request that includes a valid non-disclosure order, X does not notify the user unless permitted by law or court order to do so.

5.

a. If you notify the subscriber, how long do you wait until notification goes out?

For purposes of transparency and due process, X's policy is generally to notify users of requests for their X account information prior to disclosure of said account information.

b. Are you aware that by notifying the subscriber about a law enforcement subpoena for their subscriber information that you are jeopardizing critical evidence that could be erased before law enforcement can serve warrants?

Prior to notifying a user regarding a law enforcement request, X ensures that the requested data is preserved. Doing so, mitigates any risk of data loss.

c. Would your company agree to a 90-day non-disclosure to subscribers to allow law enforcement ample time to secure proper legal process?

To request data from X, law enforcement must seek and obtain proper legal process, thereby offering an opportunity to seek a non-disclosure order.

Do you actively seek out and incorporate feedback and insight from survivors of online sexual exploitation to improve your trust and safety policies and practices and to prevent and disrupt child sexual abuse material (CSAM) production and distribution on your platform? Can you provide examples?

We have received feedback and input from representatives of the End Online Sexual Exploitation and Abuse of Children Coalition. We welcome introductions to any other survivors or survivors groups that your office has connections to.

d. If not, please explain.

6. During our hearing, you testified that you collaborate with parents and parent organizations to create mechanisms to keep children safe online. Please elaborate and cite examples of your company's work with non-employee parents and parent organizations.

Historically, in the US, we collaborated with Parents Together on product feedback, policy, enforcement, and public policy. We commit to developing a broader set of relationships with parents and parent organizations.

7. Why does your company have the age limit of 13 years old for a user to sign up for an account?

Our age limit is in alignment with the Children's Online Privacy and Protection Act.

   a. Why not younger or older?

      To the best of my knowledge, we have not considered allowing users under 13 open accounts. In some countries internationally, the law has been set higher and we have adhered to those laws.

8. How many minors use your platform? How much money does your company make annually from these minors?

As of December 2023, in the United States, there are approximately 650,000 daily active users between 13-17.

9. What percentage of your employees work on trust and safety and how much money does your company invest annually in trust and safety?

X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.

10. It is sometimes challenging for law enforcement conducting criminal investigations to determine the true identity of a person behind a name on social media or other online platforms, and whether an online identity is an actual person. What are you doing to validate the true identity of users – or the fact that a user is a human – when they create an account on your platforms?

X uses a variety of proprietary methods to analyze account signals in order to determine the authenticity of a user. In addition, X utilizes Premium subscriptions as a method to authenticate humans because we are able to collect payment information and identification.

X uses ID verification for:

(1) User Experience Enhancement: X will provide a voluntary ID verification option for certain X features to increase the overall integrity and trust on our platform. We collect this data when X Premium subscribers optionally choose to apply for a verified badge by verifying their identity using a government-issued ID. Once confirmed, a verified label is added to the user's profile for transparency and potentially unlocking additional benefits associated with specific X features in the future. This option is currently only available to individual users and not businesses or organizations.

(2) Safety and Security Purposes: In certain instances, X may require your government-issued ID when needed to ensure the safety and security of accounts on our platform. We collect this data when investigating and enforcing our policies and may

request an ID verification in response to impersonation reports. Currently, X focuses on account authentication to prevent impersonation, and may explore additional measures, such as ensuring users have access to age-appropriate content and protecting against spam and malicious accounts, to maintain the integrity of the platform and safeguard healthy conversations.

11. Is your company using safety technology to detect and prevent live video child sexual abuse on your platforms and apps that allow users to stream or share live video? If not, please explain.

Yes, we utilize safety technology to detect and prevent child abuse in our livestream product.

a. Has your company tested that or similar technology? If not, are you developing similar technology to address child sexual abuse in live video?

We utilize technology that detects nudity and presence of children in order to detect possible CSAM livestreams. X does not allow adult content in live video.

12. How are you measuring if your trust and safety policies, practices, and tools are effective in protecting children from sexual abuse and exploitation on your platform?

We are constantly evaluating the effectiveness of our policies, tools, and enforcement on the safety of our users and the integrity of our service. We want to make X the most inhospitable place for bad actors seeking to exploit children. We evaluate the accuracy of our detection, the speed of our action, the quality of our reporting to NCMEC, the integrity of our data, the efficacy of automated interventions, and the productivity of our partnerships.

a. What specific metrics or key performance indicators do you use?

We evaluate a range of indicators, such as action rates, reporting rates, automated v. manual reporting, response rates to user reports, detection rates, to name a few.

13. Is your company using language analysis tools to detect grooming activities? If not, please explain.

a. What investments will your company make to develop new or improve existing tools?

We are currently beta testing a new text-based machine-learning classifier to detect different types of child abuse discussion, which has yielded enforcement for grooming behavior. We look forward to sharing more as we continue to develop and implement this technology.

14. What resources have you developed for victims and survivors of abuse on your platforms?

The uniqueness of X is the role it serves as a platform for public conversation, the global town square of the internet. X has always been a place for victims to bring awareness to their causes and issues of public concern, like legislation. We will continue to support organizations around the world that promote online safety and we welcome any recommendations of victims groups that we could support in their campaigns and advocacy.

15. What is your response to requests for content removal from CSAM survivors and other members of the public?

We review reports of CSE, private information, and non-consensual nudity from users and the public, as our reporting forms are available to anyone, whether or not you have an account.

16. While you mentioned several times throughout the hearing that X Corp. is a new company, it is no secret that X is Twitter rebranded. At its peak, how many trust and safety employees did Twitter have on staff and how many trust and safety employees are on staff today at X?

We have approximately 2,300 people who work on Trust & Safety and content moderation.

17. What resources does X dedicate to its child safety team and how has this team been stabilized following X's larger corporate changes over the past two years?

We have a mix of agents, policy leads, program managers, engineers, legal professionals, and government affairs professionals who work on issues of child safety, which has remained consistent over the past two years.

18. Why does X not participate in NCMEC's "Take It Down" program to help stop the sharing of and remove nude and sexually explicit photos of minors?

Our teams had recently prioritized the Tech Coalition's Project Lantern and are now evaluating the technical requirements of the program.

19. What voluntary hash-sharing or other information sharing initiatives does X participate in to help combat child sexual exploitation?

We participate in hash-sharing via NCMEC and the Technology Coalition. We have applied to the Tech Coalition's Project Lantern information sharing program. We are also evaluating participation in StopNCII.org and the NCMEC Take It Down program. We are also members of the Internet Watch Foundation and work with Thorn, in addition to international NGOs.

Senate Judiciary Committee Hearing
"Big Tech and the Online Child Sexual Exploitation Crisis"
Questions for the Record
for Linda Yaccarino
Submitted February 7, 2024

QUESTIONS FROM SENATOR SHELDON WHITEHOUSE

1. What exemptions from the protections of Section 230 would your company be willing to accept?

Section 230 is often referred to as the "26 words that made the internet" with good reason. The free, open internet would not exist in a world in which users and websites face liability for merely disseminating the speech of others. Without it, free speech on the internet would cease to exist, as websites would be forced to intensively and conservatively censor and filter speech of billions of internet users. Ideas—including ideas with the potential to improve the internet and society—would be abandoned in the face of overwhelming legal risk.

We recognize that free speech is not free, and would encourage Congress to focus on legislation that helps us hold bad actors accountable without silencing legitimate users and communities. In addition to our support for the STOP CSAM Act, we support a "bad samaritan" carve-out that would remove 230 protection from legitimately bad actors, such as websites that have the primary purpose of facilitating activities that violate federal criminal law.

2. Is it your belief that your company should enjoy absolute immunity under Section 230 from suits like *Doe v. Twitter*, No. 21-CV-00485-JCS, 2023 WL 8568911 (N.D. Cal. Dec. 11, 2023), no matter the extent of your company's failure to remove reported child sexual abuse material from the platform or to stop its distribution?

The company's legal position in the *Doe v. Twitter* matter is set forth in our public filings, and I am not in a position to comment further on the matter, as it is in active litigation. However, I note that the events at issue in that matter all took place years ago under prior management. X also supports the STOP CSAM Act, which would provide for a framework in which to hold companies civilly liable for failure to remove reported child sexual abuse material.

Linda Yaccarino – Big Tech and the Online Child Sexual Exploitation Crisis
Questions for the Record
Submitted February 7, 2024

QUESTIONS FROM SENATOR COONS

1. X Corp. ("X") has not published a transparency report with information regarding the United States since April 2023. Why has X not published any transparency report with information regarding the United States in nearly one year?

   X remains committed to transparency across the company, whether being the first amongst our peers to publish our recommendation algorithm, or making all data related to Community Notes publicly available, as well as open sourcing the code that powers it. Our goal is to return to a regular cadence of publishing global transparency reports in 2024.

2. Does X measure an estimated total amount of content on the platform that violates its suicide and self-harm policy? If not, why not?

   Yes. In 2023, more than 900,000 posts and 8,000 accounts were removed for violating our policy around promoting suicide and self harm.

   a. Does X disclose an estimated total amount of content on its platform that violates its suicide and self-harm policy? If so, please provide a specific citation to where X discloses that information. If not, why not?

      See answer to Question 1.

3. X has previously reported how much content it removes under the platform's suicide and self-harm policy.
   a. For content that has been removed, does X measure how many views that content received prior to being removed? If not, why not?

      X does measure data related to impressions for content posted on the platform.

   b. For content that has been removed, does X disclose how many views that content received prior to being removed? If so, please provide a specific citation to where X discloses that information. If not, why not?

      We continue to explore ways to share more context about how we enforce the X Rules. In a previous report, we disclosed data related to impressions of violative content. We will continue to explore the structure of future transparency reports, and we will consider data related to impressions as a reportable metric.

c. Please provide an estimate of the number of views content that was removed under this policy received in January 2024.

Data related to January 2024 will be disclosed in a future transparency report.

d. For content that has been removed, does X measure demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If not, why not?

X maintains the ability to measure demographic factors about users who view content on the platform.

e. For content that has been removed, does X disclose demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If so, please provide a specific citation to where X discloses that information. If not, why not?

X does not disclose demographic factors about users who viewed violative content. We will take this under advisement and consideration for future disclosures.

f. Does X measure the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If not, why not?

X maintains the ability to measure the number of users that have viewed violative content.

g. Does X disclose the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If so, please provide a specific citation to where X discloses that information. If not, why not?

X does not disclose data about users who viewed violative content multiple times. We will take this under advisement and consideration for future disclosures.

4. X utilizes an algorithm to recommend or amplify content to users.
   a. For content that has been removed, does X measure whether and the extent to which the removed content was recommended or amplified by X? If not, why not?

Restricting the reach of Posts, also known as visibility filtering, is one of our existing enforcement actions that allows us to move beyond the binary "leave up versus take down" approach to content moderation. However, like other social platforms, we have not historically been transparent when we have taken this action. Under certain policies, for example Hateful Conduct, we now add publicly visible labels to Posts identified as potentially violating our policies letting you know we have limited their visibility.

These labels bring a new level of transparency to enforcement actions by

displaying which policy the Post potentially violates to both the Post author and other users on X. Posts with these labels will be made less discoverable on the platform. Additionally, we will not place ads adjacent to content that we label. You can learn more about the ways we may restrict a Post's reach here. https://help.twitter.com/en/rules-and-policies/enforcement-options

b. For content that has been removed, does X disclose whether and the extent to which the removed content was recommended or amplified by X? If so, please provide a specific citation to where X discloses that information. If not, why not?

X does not disclose whether and the extent to which the removed content was recommended. We will take this under advisement and consideration for future disclosures.

c. For content that has been removed, does X measure how many views the removed content received after having been recommended or amplified? If not, why not?

X does measure data related to impressions for content posted on the platform.

d. For content that has been removed, does X disclose the number of views the removed content received after having been amplified or recommended? If so, please provide a specific citation to where X discloses that information. If not, why not?

We continue to explore ways to share more context about how we enforce the X Rules.

5. Does X support creating industry-wide transparency requirements to disclose basic safety information, like those included in the *Platform Accountability and Transparency Act*?

Yes.

Linda Yaccarino
Chief Executive Officer
X Corp.
San Francisco, CA
Questions for the Record
Submitted February 7, 2024

QUESTIONS FROM SENATOR BOOKER

1. Trust and safety teams are a vital component in combatting the spread of CSAM, hate speech, violence, and other violative content on tech platforms. Despite this, tech companies have time and time again disinvested from their trust and safety team, especially during changes in leadership.

    a. How has the size of your trust and safety team changed over the past five years? Please provide numbers for each of the past five years.

    X had 3317 Trust and Safety employees and contractors in May 2022, and 2849 in May 2023. Today, we have approximately 2300 people working on Trust and Safety matters and are building a Trust and Safety Center of Excellence in Austin, Texas, in an effort to bring more agent capacity in-house and rely less on outside contractors. We are currently hiring for full-time agent positions and have a goal of hiring approximately 100 new team members in Austin. A live job posting for this position can be found on our careers page: https://twitter.wd5.myworkdayjobs.com/X/job/Austin-TX/Agent--Trust---Safety--Content-Moderation-_R100044

    b. Do your trust and safety teams make submissions to the National Center for Missing & Exploited Children's CyberTipline, or is that a separate unit?

    Yes.

    c. If it is a separate unit, how many members are on the team and how have those numbers changed over the past five years. Please provide numbers for each of the past five years.

    N/A

2. The National Center for Missing & Exploited Children's CyberTipline plays an integral role in combatting child sexual exploitation. The tipline helps law enforcement investigate potential cases and allows prosecutors to bring justice to victims. While federal law requires your company to report to the CyberTipline any apparent violations of federal laws prohibiting child sexual abuse material of which you are aware, there are many gaps.

a. Is there a standard format your reports to the CyberTipline follow? If so, what is that format?

Yes, we will follow up with a sample CyberTipline report. Generally, a report contains a complete archive of the account, including content and media, and information such as geolocation, associated emails and/or phone numbers, and IP address(es).

b. Does your company proactively report planned or imminent offenses?

Yes.

c. Does your company proactively report potential offenses involving coercion or enticement of children?

Yes.

d. Does your company proactively report apparent child sex trafficking?

Yes.

Questions for the Record from Senator Alex Padilla
Senate Judiciary Committee
"Big Tech and the Online Child Sexual Exploitation Crisis"
Wednesday, January 31, 2024

Questions for Linda Yaccarino

1. In your testimony, you shared that less than 1% of X users are below the age of 18.
    a. How many minors are registered X account holders?

        As of January 2024, in the United States, there are approximately 650,000 daily active users between the ages of 13-17.

    b. Does X plan to provide guidance to them and their caregivers about digital online health and safety?

        X provides all users and the public information on safety and security on X via our Help Center. https://help.twitter.com/en/safety-and-security

2. In your testimony, you shared that users between the ages of 13 and 17 are automatically assigned to a private default setting and they cannot accept a message from anyone they do not approve.
    a. How are you ensuring that the burden is not on young people to make adult-level decisions about safety on the services that you operate?

        By implementing privacy and safety by default, X makes it simple for minors to use X safely.

        Accounts belonging to known minors will be defaulted to a "protected" setting. This means that known minors will receive a request when new people want to follow them (which they can approve or deny), that their posts will only be visible to their followers, and that their posts will only be searchable by them and their followers (i.e. they will not appear in public searches). Under this setting, accounts belonging to known minors will be restricted to receiving DMs from accounts they follow by default. We also utilize an age lock. Once a new user enters a date of birth that makes them under the age of 18, they will be stopped from re-entering a new date of birth for that account.

        We also take steps to limit exposure to sensitive content. Known minors or viewers who do not include a birth date on their profile are restricted from viewing specific forms of sensitive media such as adult content. X obscures sensitive media behind notices and interstitials. This includes our product age restrictions that restricts known minors from viewing adult content.

In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.

More information on our protected account settings can be found in our Help Center.
https://help.twitter.com/en/safety-and-security/public-and-protected-posts

b. In the last 4 years, how often has the company blocked products from launching because they were not safe enough for minors, or withdrawn products from the market after receiving feedback on the harms they were causing?

To the best of my knowledge, since I joined the company X has not launched products targeted at minors, nor has X withdrawn products from the market due to harms to minors.

c. Since Mr. Musk took ownership of the company, how often has the company blocked products from launching because they were not safe enough for minors, or withdrawn products from the market after receiving feedback on the harms they were causing?

To the best of my knowledge, since I joined the company X has not launched products targeted at minors, nor has X withdrawn products from the market due to harms to minors.

3. Existing detection tools for keeping child sexual abuse material from spreading online rely on hashed images of already identified CSAM imagery. There are tools like PhotoDNA and Google's CSAI match tool available for identifying this content. A challenge I hear raised frequently is identifying and removing novel images that have not already been hashed.
   a. What would it take to develop better technology to accurately identify and limit the spread of novel CSAM images?

At X we use a combination of commercial, proprietary, and open-source technology to identify, hash, report, and remove novel CSAM images, and we will continue to enhance this technology while partnering with advanced technology providers like Thorn.

X is defending itself in a class action lawsuit related to its use of one of these tools. In *Martell v. X Corp.*, Civil No. 05449 (N.D. Ill. 2023), a private plaintiff has alleged that X's use of PhotoDNA to combat CSAM violates Illinois' Biometric Information Privacy Act. Meritless lawsuits such as *Martell* discourage critical innovation in this space. We encourage Congress to explore federal protections for the development and use of anti-CSAM technologies.

b. Are there interventions from Congress that would facilitate identification of CSAM?

Investment in research and development of advanced technologies to detect CSAM would benefit the internet ecosystem.

c. Based on your company's experience trying to address online sexual exploitation and abuse of minors, are there areas where Congress could be helpful in tackling this problem?

Increased investment in law enforcement capabilities to investigate and prosecute criminals that traffic CSAM would strengthen the ecosystem as a whole. Increased investment and resources for the National Center for Missing and Exploited Children to develop their technical capabilities would accelerate the investigation and prosecution of criminals around the world.

4. AI models are making it easier to develop synthetic CSAM. These are either altered images of real people, or wholly synthetic individuals. Policymakers are grappling with what this will mean for law enforcement efforts to hold perpetrators accountable and identify children who are being harmed. In addition to processing a higher volume of Cybertips, investigators will have the added challenge of determining whether the victim in the scenario is in fact a real person. And cases are already being reported where AI generative technologies are being employed to facilitate the grooming and sextortion of minor victims.
   a. What are you doing to identify and remove AI-generated CSAM on your services?

   X utilizes a combination of technology and human capacity to enforce our Child Sexual Exploitation (CSE) policy. This policy covers media, text, illustrated, or computer-generated images.

   b. Do you flag for NCMEC if you perceive the CSAM to be AI-generated?

   Currently, we do not include this in our NCMEC reports. However, we are evaluating the requirements to include this new field in our reports.

   c. How prevalent is this kind of content?

   We do not currently have prevalence data on this type of content.

   d. How do you anticipate the rise of AI-generated CSAM will impact NCMEC's ability to process and refer Cybertips to law enforcement?

   We defer to NCMEC's expertise and experience on the impact on their ability to process and refer CyberTips to law enforcement, however, we believe that individuals who traffic computer-generated CSAM should be

investigated and prosecuted.

e. Recently, A.I.–generated explicit images of a major pop superstar were distributed widely online without her consent. That story drew attention to a growing problem over the last year facilitated by AI tools: the generation of deepfake, nonconsensual, sexually explicit imagery of everyday people, including our young people. Will you commit to reporting on the prevalence of this new problem and the steps your company is taking to address this horrendous abuse?

Non-consensual nudity (NCN) has no place on X and we will continue to rigorously enforce this policy and improve our approach. As the challenge of AI-generated NCN evolves, we will commit to working with all stakeholders to meet the challenge. We support Congressional efforts to criminalize the distribution of "deepfake" non-consensual intimate imagery.

f. Are there technical or legal barriers that your company has identified preventing thorough redteaming of AI models to ensure they do not generate CSAM?

X does not develop AI models that generate images.

5. How companies choose to allocate their resources illustrates their true priorities.
   a. What percentage of your company's budget is dedicated to addressing child safety on your platform?

Child safety is embedded into our company values and operations; there is not a specific budget allocation.

b. What process or assessment of risk on the platform informed that figure?

See answer above.

c. How many layers of leadership separates your trust and safety leaders from you and Mr. Musk?

Our Trust & Safety leadership report directly to the CEO.

d. At X, which member of the leadership team ultimately approves product decisions that impact user safety?

X's leadership across the company considers user safety in their decision-making, including product decisions.

6. The companies represented at the hearing have the money and resources to hire teams of Trust & Safety professionals and build bespoke tools to aid with content moderation and integrity work as well as the detection of content like CSAM on their services. This is not

necessarily the case for the rest of the tech sector. These are industry-wide problems and will demand industry-wide professionalization and work.

a. What is X currently doing to support access to open-source trust & safety tools for the broader tech ecosystem?

Currently, we utilize a mix of proprietary tools and licensed technologies, for example Thorn's Safer tool. (https://safer.io/) We recognize the power of open source technology and if there are technologies that can be open sourced without compromising operational integrity and preventing bad actors from exploiting the technology, we will support those efforts.

b. And if X is not doing anything now, will you commit to supporting the development of these kinds of resources?

X is committed to technology transparency, as evidenced by the publishing of our recommendation algorithm, currently available on GitHub. https://github.com/twitter/the-algorithm. The technology powering our Community Notes product is also open source, and also available on GitHub. https://github.com/twitter/communitynotes

7. One necessary element of keeping our kids safe is preventing harms in the first place. The National Center for Missing and Exploited Children partnered with the White House, the Department of Justice, and the Department of Homeland Security to create "The Safety Pledge" initiative to combat online child exploitation in September 2020. I understand more government backed public awareness campaigns are being developed.

a. Are you partnering with the federal government to distribute health and safety resources to young people?

We support the NCMEC with advertising credits to promote their campaigns on X (via their accounts @MissingKids and @NetSmartz).

b. What are you proactively doing to educate the minors that use your services about online health and safety?

We provide advertising credits to organizations around the world working on digital safety.

8. Sextortion has become increasingly prevalent. Offenders may use grooming techniques or basic trickery to manipulate victims into providing nude or partially nude images of themselves, which are then used to coerce victims into sending more graphic images and videos or pay a ransom. These criminals often threaten to post the images or sensitive images publicly or send them to the victim's friends and family if the child does not comply. From May 2022 to October 2022, U.S. law enforcement and NCMEC witnessed an alarming increase in CyberTips and reports where minors have been sextorted for money. Many young boys, including in California, have committed suicide out of desperation, leaving their loved ones devastated.

a. How is your company responding to the growing threat of financial sextortion?

We use a mixture of proactive detection, user reporting, and human capacity to enforce our private information and media policy. We are currently beta testing a text-based machine learning classifier developed by Thorn that detects sextortion, CSAM, child-access, self-generated CSAM, and child abuse discussion.

An important component in combating extortion is our private information policy. Under this policy, you may not publish or post other people's private information without their express authorization and permission. We also prohibit threatening to expose private information or incentivizing others to do so.

In addition, you may not share private media, such as images or videos of private individuals, without their consent. However, we recognise that there are instances where users may share images or videos of private individuals, who are not public figures, as part of a newsworthy event or to further public discourse on issues or events of public interest. In such cases, we may allow the media to remain on the platform.

Sharing someone's private information online without their permission, sometimes called doxxing, is a breach of their privacy and of the X Rules. Sharing private information can pose serious safety and security risks for those affected and can lead to physical, emotional, and financial hardship.

When reviewing reports under this policy, we consider a number of things, including:

- ***What type of information is being shared***
    - **We take this into consideration because certain types of private or live information carry higher risks than others, if they're shared without permission. Our primary aim is to protect individuals from potential physical harm as a result of their information being shared, so we consider information such as physical location and phone numbers to be a higher risk than other types of information. We define "live" as real-time and/or same-day information where there is potential that the individual could still be at the named location.**
- ***Who is sharing the information***
    - **We also consider who is sharing the reported information and whether or not they have the consent of the person it belongs to. We do this because we know that there are times when people may want some forms of their personal information to be shared publicly. For example, sharing a personal phone number or email for professional networking or to coordinate**

social events or publicly sharing someone's home addresses or live locations to seek help after a natural disaster.

- ***Whether the information available elsewhere online***
  - **If the reported information was shared somewhere else before it was shared on X, e.g., someone sharing their personal phone number on their own publicly accessible website, we may not treat this information as private, as the owner has made it publicly available. Note: we may take action against home addresses being shared, even if they are publicly available, due to the potential for physical harm.**
- ***Why the information is being shared***
  - **We also factor in the intent of the person sharing the information. For example, if we believe that someone is sharing information with an abusive intent, or to harass or encourage others to harass another person, we will take action. On the other hand, if someone is sharing information in an effort to help someone involved in a crisis situation like in the aftermath of a violent event, we may not take action. Note: regardless of intent, if the information is not shared during a crisis situation to assist with humanitarian efforts or in relation to public engagement events, we will remove any posts or accounts that share someone's live location.**
- ***Sharing private media***
  - **Posting images is an important part of our users' experience on X. Where individuals have a reasonable expectation of privacy in an individual piece of media, we believe they should be able to determine whether or not it is shared. Sharing such media could potentially violate users' privacy and may lead to emotional or physical harm. When we are notified by individuals depicted, or their authorized representative, that they did not consent to having media shared, we will remove the media. This policy is not applicable to public figures.**

Under this policy, you can't share the following types of private information, without the permission of the person who it belongs to:

- **home address or physical location information, including street addresses, GPS coordinates or other identifying information related to locations that are considered private;**
- **live location information, including information shared on X directly or links to 3rd-party URL(s) of travel routes, actual physical location, or other identifying information that would reveal a person's location, regardless if this information is publicly available;**
- **identity documents, including government-issued IDs and social security or other national identity numbers – note: we may make**

limited exceptions in regions where this information is not considered to be private;
- **contact information, including non-public personal phone numbers or email addresses;**
- **financial account information, including bank account and credit card details;**
- **other private information, including biometric data or medical records;**
- **media of private individuals without the permission of the person(s) depicted; and**
- **media depicting prisoners of war posted by government or state-affiliated media accounts on or after April 5, 2022.**

The following behaviors are also not permitted:
- **threatening to publicly expose someone's private information;**
- **sharing information that would enable individuals to hack or gain access to someone's private information without their consent, e.g., sharing sign-in credentials for online banking services;**
- **asking for or offering a bounty or financial reward in exchange for posting someone's private information;**
- **asking for a bounty or financial reward in exchange for not posting someone's private information, sometimes referred to as blackmail.**

b. What methods are in place to detect and disrupt this type of abuse in real time?

See above answer.

c. What kind of user education and awareness are you engaged in?

We maintain a comprehensive Help Center and regularly post on our @Safety account on X.

d. Are you aware of a higher prevalence of sexual extortion or abuse against certain demographics among young users? If not, will you commit to studying this issue and making that kind of information available to improve public education and protection measures?

We will continue to stay educated on the latest trends in exploitative behavior on our platform and across industry. Many cross-sector groups and convenings, whether NCMEC's CyberTipline Roundtable or Virtual Global Taskforce meeting, the INHOPE conference, or Tech Coalition member events, provide a great opportunity for knowledge sharing on the latest threats.

9. Young people need to be at the center of regulatory discussions, and they need to be at the table as products and services they use are designed.

a. Are you engaging young adults and youth in your conversations and policies around Trust and Safety on the platform?

We are constantly gathering input and feedback from our users around the world on issues of trust and safety.

b. How do you proactively keep up to speed with the most pressing issues facing young people online?

As the global town square of the internet, every day there is robust conversation about online safety on X, from discussion of research, to announcements about technologies or products from a range of companies, to in-depth reporting about the challenges that the internet ecosystem faces. X is the place where experts, policymakers, and parents come to discuss the most pressing issues facing young people online, and where youth activists advocate for change.

10. While employment figures alone do not reflect a company's ability to address trust & safety on their services, the precipitous decline in the trust and safety expertise and personnel at X has been alarming. According to disclosures X made to Australian regulators, between October 28, 2022, and May 31, 2023, your trust and safety staff globally had been reduced from 4,062 to 2,849 employees and contractors. Engineers focused on trust and safety issues at X had been reduced from 279 globally to 55. Full-time employee content moderators had been reduced from 107 to 51. Content moderators employed on contract fell from 2,613 to 2,305.
    a. Does X currently have a specific person in charge of Trust & Safety strategy and policy decisions?

    Yes.

    b. Who does that person directly report to?

    The CEO.

    c. How many employees at X are focused on strategic Trust & Safety matters?

    We have approximately 2300 people dedicated to Trust & Safety matters.

    d. What resources does X dedicate to child safety?

    Child safety is embedded into our company values and across all of our operations; there is not a specific budget or resource allocation.

    e. How does X's current child safety work compare financially, technically, and personnel-wise to two years ago?

An important shift and enhancement in our ability to send reports to the CyberTipline was the ten-fold increase and investment we made in training our agents in CyberTipline reporting. Prior to acquisition, we had approximately 20 agents that were trained and authorized to report to NCMEC. Now, we have approximately 200 agents that are authorized and trained and sending reports to the CyberTipline. This increase has led to an average of 64,000 manual reports per month sent to the CyberTipline, up from an average of 6,300 manual reports per month.

11. A Wall Street Journal investigation last year with the Stanford Internet Observatory and UMass Rescue Lab identified a network of more than 500 accounts of young users advertising their self-generated illicit sexual media on social media, especially Instagram and X, with tens of thousands of likely buyers. X insisted that they took action to address failures. However, months later, Stanford researchers found the problem persisted.
    a. The Stanford researchers were not able to do a complete reassessment of X because the company removed access to its Academic API offerings. Why did X retire these offerings and will you reevaluate the posture of the company with respect to academic researchers?

    We are re-evaluating our academic research program and look forward to sharing more soon.

Senator Peter Welch
Senate Judiciary Committee
Written Questions for Linda Yaccarino
Hearing on "Big Tech and the Online Child Sexual Exploitation Crisis"
January 31, 2024

X recently announced that the company plans to build a Trust and Safety Center for Excellence in Austin, Texas and will bring on 100 new employees. This announcement comes after X carried out a series of layoffs in 2023, which included cuts to X's Trust & Safety team.

1. How many people were laid off from X's Trust & Safety team in 2023?

   The Trust & Safety organization has been the least impacted by reductions in force. Since acquisition, we have right-sized the company and are hiring across departments, including Trust & Safety.

2. Why did you decide to add their jobs back? If the 100 new employees will have substantially different duties than the previously laid off employees, please describe the differences.

   Our Trust and Safety Center of Excellence will consist of a mix of agents and policy specialists, who work on a range of content moderation issues and enforcement.

3. Even before the layoffs in 2023, tech companies like X lacked appropriate policies to counter dangerous content on their platforms. How do you plan to address these issues moving forward?

   We will continue to improve the safety of the platform and our users, and we continue to conduct regular reviews of our policies, enforcement guidance, and training materials as we learn more about new content and behavioral challenges.

4. How much of your content moderation is managed by artificial intelligence?

   We use a mix of technology and human capacity to enforce our policies. Content moderation is not managed by artificial intelligence, rather by members of our Trust & Safety team.

5. Is it your view that artificial intelligence can replace human judgment in identifying and removing false or harmful content? If not, when is human judgment necessary?

   We should continue to utilize a mix of technology and human capacity to enforce our policies and maintain the integrity of our platform.

6. How have your Trust & Safety teams been trained on how to handle false or illegal AI-generated content?

X has a clear policy on the use of synthetic and manipulated media. This policy expressly provides that users "may not share synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm ('misleading media')." As part of this policy, which extends to all advertisements, X may label posts containing misleading media to help people understand their authenticity and to provide additional context.

In order for content with misleading media (including images, videos, audios, gifs, and URLs hosting relevant content) to be labeled or removed under this policy, it must:

- **Include media that is significantly and deceptively altered, manipulated, or fabricated, or**
- **Include media that is shared in a deceptive manner or with false context, and**
- **Include media likely to result in widespread confusion on public issues, impact public safety, or cause serious harm**

We use the following criteria as we consider posts and media for labeling or removal under this policy as part of our ongoing work to enforce our rules and ensure healthy and safe conversations on X:

> ***1. Is the content significantly and deceptively altered, manipulated, or fabricated?***
>
> **In order for content to be labeled or removed under this policy, we must have reason to believe that media are significantly and deceptively altered, manipulated, or fabricated. Synthetic and manipulated media take many different forms and people can employ a wide range of technologies to produce these media. Some of the factors we consider include:**
>
> - **whether media have been substantially edited or post-processed in a manner that fundamentally alters their composition, sequence, timing, or framing and distorts their meaning;**
> - **whether there are any visual or auditory information (such as new video frames, overdubbed audio, or modified subtitles) that has been added, edited, or removed that fundamentally changes the understanding, meaning, or context of the media;**
> - **whether media have been created, edited, or post-processed with enhancements or use of filters that fundamentally changes the understanding, meaning, or context of the content; and**
> - **whether media depicting a real person have been fabricated or simulated, especially through use of artificial intelligence algorithms**
>
> **We will not take action to label or remove media that have been edited in ways that do not fundamentally alter their meaning, such as retouched photos or color-corrected videos.**
>
> **In order to determine if media have been significantly and deceptively**

altered or fabricated, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been altered or fabricated, we may not take action to label or remove them.

2. Is the content shared in a deceptive manner or with false context?

We also consider whether the context in which media are shared could result in confusion or suggests a deliberate intent to deceive people about the nature or origin of the content, for example, by falsely claiming that it depicts reality. We assess the context provided alongside media to see whether it provides true and factual information. Some of the types of context we assess in order to make this determination include:

- **whether inauthentic, fictional, or produced media are presented or being endorsed as fact or reality, including produced or staged works, reenactments, or exhibitions portrayed as actual events;**
- **whether media are presented with false or misleading context surrounding the source, location, time, or authenticity of the media;**
- **whether media are presented with false or misleading context surrounding the identity of the individuals or entities visually depicted in the media;**
- **whether media are presented with misstatements or misquotations of what is being said or presented with fabricated claims of fact of what is being depicted**

We will not take action to label or remove media that have been shared with commentary or opinions that do not advance or present a misleading claim on the context of the media such as those listed above.

In order to determine if media have been shared in a deceptive manner or with false context, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been shared with false context, we will not label or remove the content.

3. Is the content likely to result in widespread confusion on public issues, impact public safety, or cause serious harm?

Posts that share misleading media are subject to removal under this policy if they are likely to cause serious harm. Some specific harms we consider include: Threats to physical safety of a person or group; incitement of abusive behavior to a person or group; risk of mass violence or widespread civil unrest; risk of impeding or complicating provision of public services, protection efforts, or emergency response. We also consider threats to the privacy or to the ability of a person or group to freely express themselves or

participate in civic events, such as: stalking or unwanted and obsessive attention; targeted content that aims to harass, intimidate, or silence someone else's voice; or voter suppression or intimidation. We also consider the time frame within which the content may be likely to impact public safety or cause serious harm, and are more likely to remove content under this policy if immediate harm is likely to result.

Posts with misleading media that are not likely to result in immediate harm but still have a potential to impact public safety, result in harm, or cause widespread confusion towards a public issue (health, environment, safety, human rights and equality, immigration, and social and political stability) may be labeled to reduce their spread and to provide additional context.

While we have other rules also intended to address these forms of harm, including our policies on violent threats, civic integrity, and hateful conduct, we will err toward removal in borderline cases that might otherwise not violate existing rules for Posts that include misleading media.

The consequences for violating our synthetic and manipulated media policy depends on the severity of the violation. For high-severity violations of the policy, including misleading media that have a serious risk of harm to individuals or communities, we will require you to remove this content. In circumstances where we do not remove content which violates this policy, we may provide additional context on posts sharing the misleading media where they appear on X. This means we may: Apply a label and/or warning message to the post; show a warning to people before they share or like the post; reduce the visibility of the post on the platform and/or prevent it from being recommended; turn off likes, replies, and Reposts; and/or provide a link to additional explanations or clarifications, such as relevant X policies. In most cases, we will take a combination of the above actions on posts we label.

If we determine that an account has advanced or continuously shares harmful misleading narratives that violate the synthetic and manipulated media policy, we may temporarily reduce the visibility of the account or lock or suspend the account. If users believe that their account was locked or suspended in error, they can submit an appeal.

7. How does X plan on addressing the large amount of disinformation that could be spread on its platform during the 2024 election?

X's purpose is to serve the public conversation, as people from around the world come together in an open and free exchange of ideas. Our team has and will continue to actively work to protect the integrity of the public conversation, by ensuring that users have access to real-time information and safeguarding the platform for everyone.

The public conversation occurring on X is never more important than during elections and other civic events. Any attempt to undermine the integrity of our service is antithetical to our fundamental rights and undermines the core tenets of freedom of expression, the value upon which our company is based.

We believe we have a responsibility to protect the integrity of all conversations from interference and manipulation. We prohibit attempts to use our services to manipulate or disrupt civic processes, including through the distribution of false or misleading information about the procedures or circumstances around participation in a civic process. This includes posting or sharing content that may suppress participation or mislead people about when, where, or how to participate in a civic process.

Beyond our policies and enforcement, we believe that X users play a pivotal role in helping provide needed context and accurate information about content that may be misleading or false.

Product Innovation: Community Notes

Community Notes, which focuses on offering context and surfacing credible information, represents a fundamental shift in how X mitigates mis- and disinformation. Community Notes aims to create a better-informed world by empowering X users to collaboratively add helpful notes to posts that might be misleading. Contributors can leave notes on any post and if enough contributors from different points of view rate that note as helpful, the note will be publicly shown the post. We believe that Community Notes is an inherently scalable and localized response to the challenge of disinformation. By making this feature an integral and highly visible part of X, and by ensuring that the user interface is simple and intuitive, we are investing in a tool that can be truly global in its application. It also reduces our reliance on forms of content moderation that are more centralized, manual and bespoke; or which require intensive and time-consuming interactions with third parties.

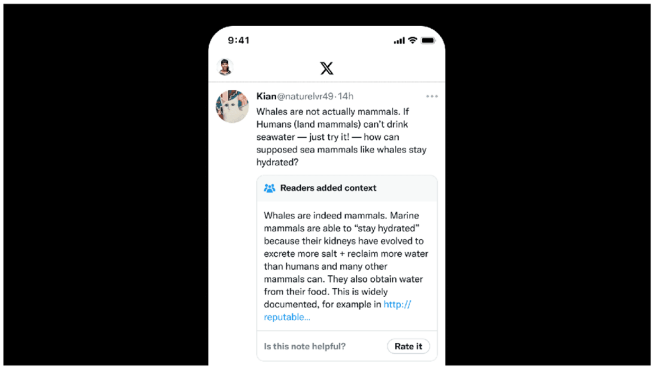
Here is how it works:

- **Contributors write and rate notes: Contributors are people on X who sign up to write and rate notes. The more people that participate, the better the program becomes.**
- **Only notes rated helpful by people from diverse perspectives appear on posts: Community Notes do not work by majority rules. To identify notes that are helpful to a wide range of people, notes require agreement between contributors who have sometimes disagreed in their past ratings. This helps prevent one-sided ratings. We have published and will continue to learn more about how Community Notes handles diverse perspectives.**
- **X does not choose what shows up, the people do: X does not write, rate, or moderate notes (unless they break the X Rules.) We believe giving people a voice to make these choices together is a fair and effective way to add**

information that helps people stay better informed.

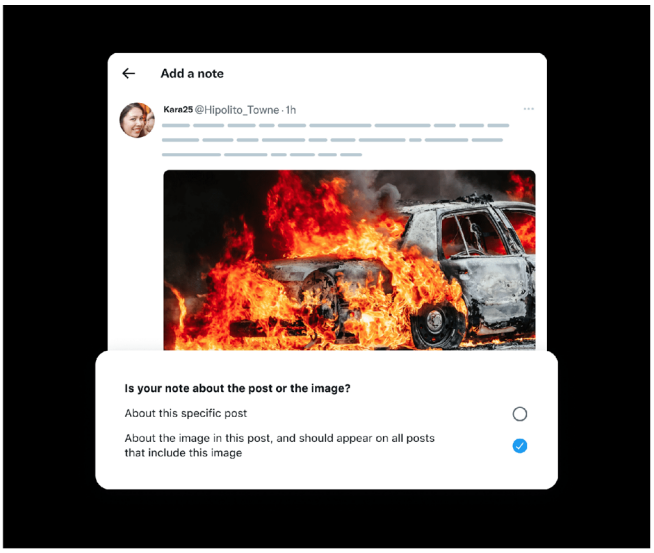
- **Open-source and transparent: It is important for people to understand how Community Notes work to be able to help shape it.**
- **The program is built on transparency: all contributions are published daily, and our ranking algorithm can be inspected by anyone. Learn more about how it works through our dedicated Community Notes Guide.**

We acknowledge and are keenly aware that a product like this can be subject to attempts of abuse and manipulation, which we proactively assess. You can read more here on how we are thinking about quality control, guardrails, circuit breakers, and the various remediations we have in place to challenge bad actors.



Notes on Media

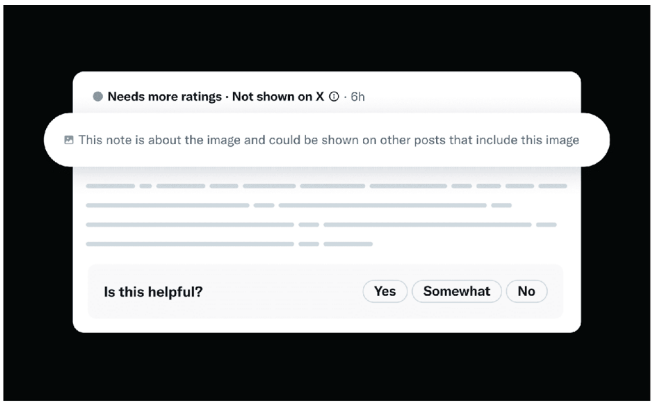
Community Notes are frequently added to posts that feature images or videos. In many cases, these notes can provide valuable context, not just for a single post, but for any post containing the same media. This feature is especially important for addressing the challenges of media produced by generative artificial intelligence tools. Contributors with a Writing Impact score of 10 or above have the option to write notes about the media found within posts, as opposed to focusing on the specific post. Contributors should select this option when they believe the context added would be helpful independently of the post the note is attached to.



Tagging notes as "about the image" makes them visible on all posts that our system identifies as containing the same image. These notes, when deemed Helpful, accumulate view counts from all the posts they appear in, but only count as one Writing and Rating Impact for the author and raters.

When someone rates a media note, the rating is associated with the post on which the note appeared. This allows Community Notes to identify cases where a note may not apply to a specific post.

For example, because of this new feature, notes related to the Israel-Hamas conflict have been displayed on 10,000+ posts. This number grows automatically if the relevant images and video are re-used in new posts.



Currently, this feature is experimental and only supports posts with a single image. We are actively working on expanding it to support posts with multiple images, GIFs, and videos. Stay tuned for updates.

Effectiveness & Research

We already know that the Community Notes feature is effective. According to the results of four surveys run at different times between August, 2021 and August, 2022, a person who sees a Community Note is, on average, 20-40% less likely to agree with the substance of a potentially misleading post than someone who sees the post alone. Survey participation ranged from 3,000 to more than 19,000 participants, and the results were consistent throughout the course of the year, even as news and post topics changed. We also see that Community Notes informs sharing behavior. Analyzing our internal data, we have found that a person on X who sees a note is, on average, 15-35% less likely to like or repost a post than someone who sees the post alone. In our most recent survey, Community Notes were found to be informative regardless of a person's self-identified political party affiliation — there was no statistically significant difference in average informativeness across party identification.

We published a research paper on Community Notes that you can read here. It goes into more detail on how we have been measuring efficacy. In addition, all Community Notes contributions are publicly available on the Download Data page of the Community Notes site so that anyone has free access to analyze the data, identify problems, and spot opportunities to make the product better.

We know there are many challenges involved in building an open, participatory system like Community Notes — from making it resistant to manipulation attempts, to ensuring it is not dominated by a simple majority or biased because of the distribution of contributors.

We have been building Community Notes (formerly called Birdwatch) in public since January 2021, and have intentionally designed it to mitigate potential risks. We have seen encouraging results, but we are constantly designing for challenges that could arise. Here are a handful of particular challenges we are aware of as well as steps we are taking to address them:

Preventing Coordinated Manipulation Attempts

Attempts at coordinated manipulation represent a crucial risk for open rating systems. We expect such attempts to occur, and for Community Notes to be effective, it needs to be resistant to them. The program currently takes multiple steps to reduce the potential for this type of manipulation:

- **First, all X accounts must meet the eligibility criteria to become a Community Notes contributor. For example, having a unique, verified phone number. These criteria are designed to help prevent the creation of large numbers of fake or sock puppet contributor accounts that could be used for inauthentic rating.**
- **Second, Community Notes does not work like many engagement-based ranking systems, where popular content gains the most visibility and people can coordinate to mass upvote or downvote content they do not like or agree with. Instead, Community Notes uses a bridging algorithm — for a note to be shown on a post, it needs to be found helpful by people who have tended to disagree in their past ratings. Academic research indicates that bridging-based ranking can help to identify content that is healthier and higher quality, and reduce the risk of elevating polarizing content.**
- **In addition to requiring ratings from a diversity of contributors, Community Notes has a reputation system in which contributors earn helpfulness scores for contributions that people from a wide range of perspectives find helpful. Helpfulness scores give more influence to people with a track record of making high-quality contributions to Community Notes, and lower influence to new accounts that have yet to demonstrate a track record of helpful ratings and contributions.**
- **Lastly, Community Notes tracks metrics that alert the team if suspicious activity is detected, and has a set of guardrails and procedures to identify if contribution quality falls below set thresholds. This helps Community Notes to proactively detect potential coordination attempts and impacts to note quality.**

Reflecting Diverse Perspectives, Avoiding Biased Outcomes

Community Notes will be most effective if the context it produces can be found to be helpful by people of multiple points of view and not just people from one group or another. To work towards this goal, Community Notes currently takes the following steps:

- **First, as described above, Community Notes uses a bridging based algorithm**

to identify notes that are likely to be helpful to people from many points of view. This helps to prevent one-sided ratings and to prevent a single group from being able to engage in mass voting to determine what notes are shown.

- **Second, Community Notes can proactively seek ratings from contributors who are likely to provide a different perspective based on their rating history. This is currently done in the Needs Your Help tab, and we are exploring new ways to quickly collect ratings on notes from a wide range of contributors.**
- **Third, to help ensure that people of diverse backgrounds and viewpoints feel safe and empowered to participate, Community Notes has implemented program aliases that are not publicly associated with contributors' X accounts. This can help prevent one-sidedness by providing more diverse contributors with a voice in the system.**
- **Finally, we regularly survey representative samples of X customers who are not Community Notes contributors to assess whether a broad range of people on X are likely to find the context in Community Notes to be helpful, and whether the notes can be informative to people of different points of view. This is one indicator of Community Notes' ability to be of value to people from a wide range of perspectives vs. to be biased towards one group or viewpoint. X customers who are not enrolled Community Notes contributors can also provide rating feedback on notes they see on X. This provides an additional indicator of note helpfulness observed over time.**

Expansion & Localization

Community Notes are now publicly visible to everyone on X. Users in 65 countries, including the US, the UK, Japan, Brazil, Mexico, Ireland, Canada, Spain, Portugal, Italy, Germany, Austria, Belgium, France, Switzerland, Luxemburg, Netherlands, Australia, New Zealand, Slovakia, Algeria, Bahrain, Egypt, Israel, Jordan, Kuwait, Lebanon, Morocco, Oman, Palestinian Territories, Qatar, Tunisia, United Arab Emirates, and just recently, Hong Kong, South Korea, and Taiwan, can now contribute to the program. Over the coming months, users in more markets will be able to contribute notes and the product will be localized further. We currently have over 400,000 users enrolled in Community Notes.

Over time, users writing in any language should be able to contribute to Community Notes and the most helpful contributions will be surfaced to inform readers. Eventually, we can see a future where attempts to spread disinformation are consistently flagged by conscientious users seeking to share important context and facts with citations.

The technology-first strategy evidenced by Community Notes is reflective of how we intend to approach content moderation going forward. We believe that this approach has obvious advantages over more centralized methods of content moderation, which have always faced the same two challenges: speed and scale.

This is an open and transparent process. That is why we have made the Community

Notes algorithm open source and publicly available on GitHub, along with the data that powers it so anyone can audit, analyze or suggest improvements.

X recently instituted full end-to-end encryption for some direct messages on its platform. All questions below pertain to X's use of end-to-end encryption.

8. Please describe why you chose to implement end-to-end encryption for direct messages and the benefits you see from it.

   At this time, X does not offer full end-to-end encryption. Encrypted Direct Messages are limited to only subscribers of X Premium, and only text is encrypted.

   Users need to satisfy the following conditions in order to send and receive encrypted messages:
   a. **both sender and recipient are on the latest X apps (iOS, Android, Web);**
   b. **both sender and recipient are verified users or affiliates to a verified organization; and**
   c. **the recipient follows the sender, or has sent a message to sender previously, or has accepted a Direct Message request from the sender before.**

9. How do you balance the benefits of encryption with the need for law enforcement to be able to track down wrongdoers on your platform?

   We limit the availability of encryption to X Premium subscribers, which reduces the number of accounts eligible to use encryption, while also providing more personal and financial information that law enforcement could utilize.

10. Given X's recent implementation of end-to-end encryption, please explain the steps and processes you use to identify child sexual abuse material, how you remove it, and how you report it to law enforcement.

    By encrypting only text, we are able to scan Direct Messages for known or potential CSAM.

    Anyone can report potential CSAM, whether they have an X account or not. In the majority of cases, the consequence for violating our CSE and CSAM policy is immediate and permanent suspension from the platform. In addition, violators will be prohibited from creating any new accounts in the future. When we are made aware of content depicting or promoting child sexual exploitation, including links to third party sites where this content can be accessed, we immediately remove it without further notice and report to the National Center for Missing & Exploited Children (NCMEC).

    Our proactive detection efforts are primarily driven by the following tools:
    - **Hash-sharing: Our current methods of surfacing potentially violating content for human review include leveraging the hashes provided by NCMEC and**

industry partners - which makes this the most widely used form of CSAM detection. We scan all media uploaded to X for matches to hashes of known CSAM sourced from NGOs, law enforcement, and other platforms. Users posting known content are immediately permanently suspended and reported to NCMEC. For videos we use a proprietary hashing algorithm produced by Thorn.

- **Automatic text detection: We have a variety of tools to access the likelihood that a post is advertising or promoting the sharing of child sexual abuse material. Some of these defenses lead to automatic suspensions while other users are flagged for human review.**
- **For videos we use a proprietary hashing algorithm produced by Thorn.**
- **PhotoDNA and internal proprietary tools: a combination of technology solutions are used to surface accounts violating our rules on Child Sexual Exploitation.**
- **Media Risk Scanning: We receive a media classifier score through Safer and it is used to filter false positive hash matches at the moment. We use a novel classifier model to rate media shared through posts' likelihood of being CSAM. Media that receives a high score is then sent to human review.**

Additionally, any attempts to circumvent an enforcement action (such as a permanent suspension) by creating additional accounts or repurposing existing accounts to replace or mimic a suspended account are considered a violation of our ban evasion policy and it will result in permanent suspension at first detection.

Post-Hearing Questions for the Record
Submitted to X CEO Linda Yaccarino
From Senator Laphonza Butler

"Protecting Our Children Online: Big Tech and the Crisis of Online Child Sexual Exploitation"

January 31, 2024

1. **Family and parental control tools:** I was glad to hear that you have spent time talking with parents and what their families need from your products. I was also glad to hear your companies have a Family Center, or other similar tools, to give parents more insight and control over how their children are using your platforms and apps.

   a. How do you advertise this feature to parents?

      We do not currently have parental tools or a Family Center.

   b. Can you share data on how many Family Center/parental tools users there are in proportion to total minors on your platforms and products?

      N/A

a.

Questions for Linda Yaccarino (X) - Grassley

Please answer each question to the fullest possible extent. If your platform is unable to answer a particular question or does not have requested data, explain why. Each question refers to your company in addition to any corporate affiliates, including parent and subsidiary companies.

1. Current law requires that a provider of a report of suspected CSAM to the National Center for Missing and Exploited Children's (NCMEC) CyberTipline preserve "any visual depictions, data, or other digital files that are reasonably accessible and may provide context or additional information about the reported material or person" for a minimum of 90 days. 18 U.S.C. 2258A(h)(1-2). The recent explosion of suspected abuse has presented unprecedented challenges for law enforcement to follow up on leads before companies discard or delete essential data and information. There is nothing preventing tech companies from preserving relevant material beyond the statutorily-mandated 90-day period.
    a. How long does X voluntarily preserve and retain data contained in and related to its reports to the CyberTipline?

    X retains the associated account data for 90 days, unless it is subject to a legal obligation to preserve the data for a longer period. However, this data is provided to NCMEC who may then retain the data for investigative purposes in line with their own retention policies.

    b. The massive influx of reports to the CyberTipline naturally results in law enforcement entities having to conduct and finish investigations beyond 90 days of an initial report to the CyberTipline. Retaining relevant information for longer periods could significantly advance law enforcement's ability to thoroughly investigate leads. If X only preserves and retains this information for the minimum 90-day period, why does it do so when preserving this data longer could significantly enhance and prolong law enforcement's ability to investigate and prosecute child predators?

    As described above, NCMEC, in its role as the primary interlocutor with law enforcement, is able to preserve the data provided by X as long as it deems necessary and in line with their retention policies. X has no objections to retaining the associated account data for longer than the current 90 day period should such an extension prove useful to law enforcement. This also highlights the importance of law enforcement having sufficient resources to investigate reports promptly.

    c. Please confirm if X stores and retains the following information relating to reports to the CyberTipline:
        i. IP addresses - **YES**
        ii. Screen Names - **YES**

iii. User Profiles - **YES**
iv. Associated Screennames (by IP address and associated emails) - **YES**
v. Email addresses - **YES**
vi. Geolocation data - **YES**

d. If X does not retain or store any of the above types of information in question (c), please explain why.

N/A

e. Please list any other information X retains and preserves for law enforcement purposes not listed above in question (c).

Any content and media associated with the CyberTipline report, including an entire archive of the user's account.

f. Does X flag screennames and associated email addresses to suspected accounts that violate X's terms of service?

Yes.

2. How does X prioritize urgent requests for information from law enforcement and what is X's response time to urgent requests?

X has a dedicated online portal for emergency information requests submitted by law enforcement. This portal is monitored 24/7 and all submissions are handled on a priority basis with responses in less than two hours.

3. What is X's average response time to service of legal process from law enforcement for CSAM-related information?

X endeavors to respond to legal process received from law enforcement and appropriate government entities in a prompt manner. Specifically, X endeavors to respond prior to the enumerated production date required by law in the particular jurisdiction or outlined in the individual legal process.
Upon identifying that a request seeks CSAM-related information, X prioritizes and handles such requests on an expedited basis.

In 2023, the tech industry as a whole slashed more than 260,000 jobs. And in the first four weeks of this year, another 25,000 jobs were cut.

a. For each year, between 2018 and 2023, how many U.S. based employees did you have at X?

2018 - Approximately 3920 employees globally
2019 - Approximately 4900 employees globally
2020 - Approximately 5500 employees globally
2021 - Approximately 7500 employees globally
2022 - Approximately 8000 employees globally (through Q3)
2023 - Approximately 1500 employees globally

We are continuing to investigate the breakdown of US based employees and will follow up with your staff.

i. Of these employees, how many were sponsored on H-1B visas?

In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

ii. For each year, between 2018 and 2023, how many H1-B visa applications did X submit?

In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

b. For each year, between 2018 and 2023, how many employees based outside the U.S. did you have at X?

In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

i. Of these employees, how many were based in China?

The company does not have operations in China.

c. For each year, between 2018 and 2023, how many employees in total did X terminate, fire, or lay off?

i. Of these employees, how many were based in the United States?

In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

ii. Did X fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?

In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

iii. Were any duties and/or functions previously performed by laid-off employees transferred to or performed at any point by employees

sponsored on H1-B visas? If so, which duties and/or functions?

In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

d. For each year, between 2018 and 2023, how many employees performing work related to child safety did X terminate, fire, or lay off?
   i. Of these employees, how many were based in the United States?

   In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

   ii. Did X fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?

   In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

   iii. Were any duties and/or functions (specifically relating to child safety) previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?

   In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.

   iv. How have layoffs impacted X's ability to protect children on its platforms?

   As evidenced by the data on our enforcement, increased manual and automated reporting to NCMEC, implementation of advanced technologies, ten-fold increase in training of agents, and speed of execution on our number one priority, our capacity to combat CSAM has not been impacted, rather enhanced.

   v. Does X have any plans to increase staff responsible for child safety operations or otherwise optimize its child safety operations?

   We are building a Trust & Safety Center of Excellence in Austin, Texas, with a goal of hiring 100 full-time agents that will contribute to our safety operations.

4. On January 30, 2024, the Tech Transparency Project (TTP) published an article on their website called, "Meta Approves Harmful Teen Ads with Images from its Own AI Tool". In summary, TTP, using Meta's "Imagine with Meta AI" tool generated inappropriate images such as young people at a pill party or other vaping. These images with text were

submitted to Facebook as advertisements targeting users between ages 13-17 in the United States. TTP reported that Facebook approved the advertisement, despite it violating its own policies, in less than five minutes to run on the following platforms: Facebook, Instagram, Messenger, and Meta Quest. Meta. Over the course of a week, TTP submitted the advertisements with the same end result: Facebook approving them. TTP reported that they canceled these advertisements before their scheduled publication, but it illustrated the repeated failures of Facebook to properly moderate content. This is just one example of what other non-government organizations and others have uncovered across social media platforms.

a. How often a month do X employees conduct quality checks on X's policies and safeguards for child accounts?

We are constantly evaluating our systems and protocols for ensuring the safety of our product. We do not allow minors under 13 to have accounts.

b. In which departments, components, or units of the company does X have staff dedicated to performing this type of work?

Quality assurance and integrity of our work is built into our operational ethos and implemented across the company.

c. How many employees make up these departments, components, or units?

Safety of minors is a shared responsibility across the entire company.

d. If a violation is found, what action is taken, and how quickly is action taken?

5. Social media companies claim they are investing in company components dedicated to safety, and that their platforms are safe for children. However, children continue to be exploited daily across these platforms.

a. What have X's revenue and profit figures been for the last three years (2021-2023)? Please provide figures broken out per year. Do not provide percentages.

Revenue for the years 2021 and first two quarters of 2022 were disclosed as part of Twitter's public company filings.

2021 revenue was $5.08 billion, with a 2021 operating loss of $493 million and net loss of $221 million.

2022 Q1 revenue was $1.2 billion, an operating loss of $128 million, and a net income of $513 million which includes a pre-tax gain of $970 million from the sale of MoPub for $1.05 billion and income taxes related to the gain of $331 million.

2022 Q2 revenue was $1.18 billion, with an operating loss of $344 million, and a net loss of $270 million.

Beginning in Q3 2022, X Corp., a private company, took over operation of the platform. As a privately-held company, X does not maintain or release public financial statements.

b. How much has X spent in advertising for the last three years (2021-2023), broken out per year?

As a public company in 2021, Twitter disclosed its annual financial statements as part of its public filings to the SEC. This mandatory disclosure included expenditures related to 'Sales and marketing'. Sales and marketing expenses consist primarily of personnel-related costs, including salaries, commissions, benefits and stock-based compensation for our employees engaged in sales, sales support, business development and media, marketing, corporate communications and customer service functions. In addition, marketing and sales-related expenses also include advertising costs, market research, trade shows, branding, marketing, public relations costs, amortization of acquired intangible assets, allocated facilities costs, and other supporting overhead costs. In its last annual disclosure, sales and marketing expenditures totaled $1,175,970,000.

Beginning in Q3 2022, X Corp., a private company, took over operation of the platform. X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.

c. How much of X's resources spent on advertising has been devoted to advertising X's safety initiatives and efforts for the last three years (2021-2023), broken out per year?

X does not have a specific budget line item for advertising safety initiatives.

d. To get an understanding of how your company has invested and plans to invest in its components dedicated to child safety functions, what are the annual budgets for X's child safety-related components for the last three years (2021-2023)?

X does not have a specific budget line item for 'child safety-related components,' as child safety is not limited to specific components and is prioritized across the entire Trust & Safety organization.

e. What is the current anticipated (2024) budget for X's child safety-related components?

X does not have a specific budget line item for 'child safety-related components,' as child safety is not limited to specific components and is

prioritized across the entire Trust & Safety organization.

f. Provide the number of staff employed in X's child safety-related components for the last three years (2021-2023).

X does not allocate a specific number of staff for child safety-related components.

g. How much is that compared to X's other components for the same period? (Please provide a breakout per year. Do not provide percentages.)

Our Trust & Safety teams are cross-functional and work across a variety of issues.

h. How many staff are currently employed in X's child safety-related components?

X does not allocate a specific number of staff for child safety-related components.

i. What are the roles, responsibilities, and functions of X's child safety-related components?

X has a combination of program managers, policy specialists, operations specialists, engineers, legal professionals, government affairs professionals, and other functions that work on issues related to child safety.

j. Are any other components responsible for the monitoring of CSAM on X's platform(s)?

See answer (i).

k. What, if any, third parties does X employ or contract with to address CSAM material on its platforms?

X partners with a network of vendors that supplement our Trust & Safety agent capacity.

i. What are the roles and responsibilities of these third parties?

Generally, the external Trust & Safety agent workforce conducts reviews of user reports, escalated and surfaced content, enforce our platform policies, and resolve account issues. This agent workforce undergoes regular and intensive training on our policies and enforcement guidance.

ii. What is the breakdown of cost per third party over the last three years (2021-2023)?

At this time we are not able to provide a specific cost breakdown of our partnerships with the network of Trust & Safety vendors.

6. Of all reports sent by X to the National Center for Missing and Exploited Children, how many reports were self-generated from victim users for the last three years (2021-2023)? Please provide the actual number of self-generated reports in addition to the total number of reports (including those that were not self-generated). In addition, please provide a break-down of the self-reporters by age.

X does not capture whether the report is self-generated from victim users.

7. What is X's policy or protocol with respect to law enforcement accessing user data and subsequent notification to users of law enforcement accessing their data?

Non-public information about X users will not be released to law enforcement except in response to appropriate legal process such as a subpoena, court order, other valid legal process, or in response to a valid emergency request.

Requests for the contents of communications (e.g., posts or photos) require a valid search warrant or equivalent from an agency with proper jurisdiction over X.

For purposes of transparency and due process, X's policy is to notify users (e.g., prior to disclosure of account information) of requests for their X account information, including a copy of the request, unless we are prohibited from doing so (e.g., an order under 18 U.S.C. § 2705(b)) or an exception applies (e.g., imminent threat to life, child sexual exploitation, or terrorism). We may ask that any non-disclosure provisions include a specified duration (e.g., 90 days) during which X is prohibited from notifying the user or may object to the non-disclosure order on free speech or other grounds.

a. Do certain crimes such as drug trafficking or child exploitation affect X's decision to notify a user whose data is accessed by law enforcement?

Yes. Exceptions to our user notice policy may include exigent circumstances or circumstances where our commitment to user notice is outweighed by the negative effects such notice would have on law enforcement efforts, such as emergencies regarding imminent threat to life, child sexual exploitation, or terrorism.

b. Do certain requests such as a subpoena or search warrant affect X's notification protocol? If so, what are they?

No, user notification rests on whether there is a non-disclosure provision in

the legal process provided by law enforcement and the presence of a policy exception.

c. If X does notify users of law enforcement accessing their data, why does X find this necessary?

For purposes of transparency and due process, X's policy is generally to notify users of requests for their X account information prior to disclosure of said account information.

8. The National Center for Missing and Exploited Children has indicated that reports from social media companies tend to lack actionable information causing law enforcement to be burdened with incomplete information. How comprehensive are X's reports to NCMEC? What challenges is X experiencing on the collection of user data and other information to include in its reports to NCMEC? What actions is X taking to make its reports more comprehensive and useful to law enforcement?

When we report a user to NCMEC, we include a full archive of that user's data for law enforcement to access via a subpoena for their investigation. From what NCMEC has shared with us, we provide significantly more user data than most companies, which helps law enforcement prosecute violators. We are working on a variety of changes to our system to increase data throughput and reduce NCMEC submission error rates.

Senator Mike Lee
Questions for the Record
Linda Yaccarino, Chief Executive Officer, X Corp.
Hearing on "Big Tech and the Online Child Exploitation Crisis"
Submitted February 7, 2024

1. In last week's hearing, you stated that "[i]t's time to criminalize the sharing of nonconsensual material." Will you support the PROTECT Act of 2024, S. 3718, which creates a pathway for victims of nonconsensual image dissemination to easily have their images removed and creates civil and criminal liability for platforms who knowingly share those nonconsensual images?

   We support the SHIELD Act, and we will evaluate the PROTECT Act and look forward to discussing it with you and your staff.

2. The 2022 Thorn Report indicated that 19 percent of minors who use X have had an online sexual interaction on your platform, and 13 percent of minors who use X had an online sexual interaction with someone they believed to be an adult. What are you doing to put an end to these interactions?

   In early 2023, after the acquisition, we implemented safety-by-default features for accounts opened by users between the ages of 13-17. Accounts belonging to known minors are now defaulted to a "protected" setting. This means that known minors will receive a request when new people want to follow them (which they can approve or deny), that their posts will only be visible to their followers, and that their posts will only be searchable by them and their followers (i.e. they will not appear in public searches). Under this setting, accounts belonging to known minors will be restricted to receiving DMs from accounts they follow by default. We also utilize an age lock. Once a new user enters a date of birth that makes them under the age of 18, they will be stopped from re-entering a new date of birth for that account.

   We also take steps to limit exposure to sensitive content. Known minors or viewers who do not include a birth date on their profile are restricted from viewing specific forms of sensitive media such as adult content. X obscures sensitive media behind notices and interstitials. This includes our product age restrictions that restricts known minors from viewing adult content.

   In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.

In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.

More information on our protected account settings can be found in our Help Center.
https://help.twitter.com/en/safety-and-security/public-and-protected-posts

3. X's policies permit pornography on your platform. An estimated 13 percent of total material on X is explicit. How do you guarantee that minors cannot see any of this material on your platform?

X utilizes content filters for sensitive media and minor account holders are restricted from seeing content marked as sensitive. X may also use automated techniques to detect and label potentially sensitive media, and to detect and label accounts that frequently post potentially sensitive media.

4. X restricts certain content from accounts that belong to minors. However, the only age verification measure that X undertakes to ascertain the age of its users is asking new users to enter their birthdate when they open an account. How do you prevent minors from lying about their age when creating an account?

X's age assurance process relies on self-declaration to collect the user's date of birth through the neutral presentation of a date of birth prompt and allows any user to report other users who they believe are under the age of 13. X has set up a dedicated age moderation workflow to enable any user to report an account that they suspect is being used by a minor under the age of 13, available at https://help.twitter.com/en/forms/safety-and-sensitive-content/underage-user.

X strongly supports app stores handling age-gating for apps. Age verification through app stores is simply leveraging existing processes, enhances teen safety by filtering all inappropriate apps and preserves privacy by avoiding the need for personal information sharing at the individual app level.

5. Currently, X only encrypts private messages for users with premium accounts. If a minor upgrades to a premium account, will you encrypt their messages as well? Will this hamper your ability to detect sexual exploitation of minors or grooming behaviors?

Premium subscriptions require credit card information, which minors should not have unless authorized by a parent. Direct Messages are not encrypted by default, rather premium subscribers must opt into a new encrypted message, and only text is encrypted, which allows us to continue scanning DM surfaces for violative content and CSAM.

6. In 2021, there was a widely-reported failure by Twitter to suspend accounts sharing CSAM and remove nonconsensual videos of a 13-year-old minor that were shared on your platform. The minor had been a victim of sextortion, and after he had ceased communicating with the predators, they shared his images on Twitter. The minor became aware of his images which appeared on two well-known CSAM Twitter accounts and had been viewed nearly 200,000 times. The minor made three separate requests to Twitter to have his images removed—including one in which he sent a copy of his ID to verify that it was him in the videos. Despite these efforts, the minor was told that in Twitter's review of those videos and those accounts, you "didn't find a violation of our policies." Twitter only acted on those videos and accounts after a federal agent made a "take-down demand." This event happened prior to Twitter's acquisition by Elon Musk and the shift in Twitter's, now X's, priorities. However, in May of 2023, researchers at the Stanford Internet Observatory marked 40 images of CSAM that remained on X's platform over a period of two months, despite those images being marked at CSAM by Microsoft's PhotoDNA tool and existing in NCMEC's photo-hashing database. What is the current process for a person to have their nonconsensual images removed from X? What is the maximum amount of time you permit those images to remain before removal? What changes have you made since 2021 to ensure that the 13-year-old minor's experience will never be repeated?

X maintains a policy against the sharing of non-consensual intimate media.

You may not post or share intimate photos or videos of someone that were produced or distributed without their consent.

Sharing explicit sexual images or videos of someone online without their consent is a severe violation of their privacy and the X Rules. Sometimes referred to as revenge porn, this content poses serious safety and security risks for people affected and can lead to physical, emotional, and financial hardship.

Under this policy, you can't post or share explicit images or videos

that were taken, appear to have been taken or that were shared without the consent of the people involved.

Examples of the types of content that violate this policy include, but are not limited to:

- **hidden camera content featuring nudity, partial nudity, and/or sexual acts;**
- **creepshots or upskirts - images or videos taken of people's buttocks, up an individual's skirt/dress or other clothes that allows people to see the person's genitals, buttocks, or breasts;**
- **images or videos that superimpose or otherwise digitally manipulate an individual's face onto another person's nude body;**
- **images or videos that are taken in an intimate setting and not intended for public distribution; and**
- **offering a bounty or financial reward in exchange for intimate images or videos.**

We will immediately and permanently suspend any account that we identify as the original poster of intimate media that was created or shared without consent. We will do the same with any account that posts only this type of content, e.g., accounts dedicated to sharing upskirt images.

In other cases, we may not suspend an account immediately. This is because some people share this content inadvertently, to express shock, disbelief or to denounce this practice. In these cases, we will require you to remove this content. We will also temporarily lock you out of your account before you can post again. If you violate this policy again after your first warning, your account will be permanently suspended.

We are also evaluating the technical requirements to participate in NCMEC's Take It Down program.

7. What is X doing to prohibit known CSAM from being uploaded and shared on your platform?

Media hash matching is one of the ways we take down instances of known CSAM circulating on the platform. In December 2022, we launched a new hash matching pipeline called Safer through our partnership with Thorn that allows us to take down more media than before. Through Thorn we also have access to a CSAM media classifier for the first time, which allows us to detect previously unseen images. In 2023, we detected over 60,000 pieces of media through Safer.

We also leverage hash databases that are maintained by NCMEC and the Tech Coalition to detect known CSAM on X. We scan all media uploaded to X for matches to hashes of known CSAM sourced from NGOs, law enforcement and other platforms. Users posting known content are immediately permanently suspended and reported to NCMEC.

8. Predators often engage with a minor and then quickly attempt to move the conversation to another app with more encryption protections. What does X do to identify these types of interactions, and what does X do to prevent them? Do you report to NCMEC when you suspect that a predator is engaged in grooming a minor?

Yes, we do report this behavior to NCMEC. Also, we recently submitted our application to join Project Lantern, a program developed by the Tech Coalition. The purpose of this initiative is to enable tech and related industry companies to share information and signals to root out cross-platform bad actors.

9. Do you provide information provided to NCMEC regarding suspected grooming or sexual abuse to a minor's parents?

No, as we do not collect information of parents as they are not linked to accounts of minors.

SENATOR TED CRUZ

U.S. Senate Committee on the Judiciary

Questions for the Record for Linda Yaccarino, CEO, X

I. Directions

Please provide a wholly contained answer to each question. A question's answer should not cross-reference answers provided in other questions.

If a question asks for a yes or no answer, please provide a yes or no answer first and then provide subsequent explanation. If the answer to a yes or no question is sometimes yes and sometimes no, please state such first and then describe the circumstances giving rise to each answer.

If a question asks for a choice between two options, please begin by stating which option applies, or both, or neither, followed by any subsequent explanation.

If you disagree with the premise of a question, please answer the question as-written and then articulate both the premise about which you disagree and the basis for that disagreement.

If you lack a basis for knowing the answer to a question, please first describe what efforts you have taken to ascertain an answer to the question and then provide your tentative answer as a consequence of its reasonable investigation. If even a tentative answer is impossible at this time, please state why such an answer is impossible and what efforts you intend to take to provide an answer in the future. Please further give an estimate as to when Senator Cruz will receive that answer.

To the extent that an answer depends on an ambiguity in the question asked, please state the ambiguity you perceive in the question, and provide multiple answers which articulate each possible reasonable interpretation of the question in light of the ambiguity.

II. Questions

1. In the last two years, has an employee or commissioner of the Federal Trade Commission (FTC) requested to evaluate or evaluated your data used for training Large Language Models or algorithms or the sources of such data for bias, discrimination, or misinformation?

   No.

2. In the last two years, has an employee or commissioner of the FTC sought details regarding your company's measures related to filtering or blocking inputs and outputs of a Large Language Model or algorithms.

   a. If yes, has the FTC attempted to coerce or otherwise request you to implement input/output filtering in order to allegedly comply with federal law?

   No.

3. In the last two years, has an employee or commissioner of the Federal Trade Commission sought to evaluate your company's use of measures, including "prebunking" or "debunking", designed to counteract so called "online misinformation"?

   No.

4. In June 2022, the FTC released a report titled "Combatting Online Harms Through Innovation." In this report, the FTC discussed how the deployment of AI tools intended to detect or otherwise address harmful online content is accelerating but may never be appropriate as an alternative to human judgment.

   a. In the context of protecting children from online harms to what extent does your company rely on automated tools to detect online harm vs. human review? Please be specific.

      In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year. For the first time ever, we are evaluating all videos and GIFs posted on X for CSAM. Since

launching this new approach in July 2023, we have matched over 70,000 pieces of media.

Our proactive detection efforts are primarily driven by the following tools:

- **Hash-sharing: Our current methods of surfacing potentially violating content for human review include leveraging the hashes provided by NCMEC and industry partners - which makes this the most widely used form of CSAM detection. We scan all media uploaded to X for matches to hashes of known CSAM sourced from NGOs, law enforcement and other platforms. Users posting known content are immediately permanently suspended and reported to NCMEC. For videos we use a proprietary hashing algorithm produced by Thorn.**
- **Automatic text detection: We have a variety of tools to access the likelihood that a post is advertising or promoting the sharing of child sexual abuse material. Some of these defenses lead to automatic suspensions while other users are flagged for human review.**
- **For videos we use a proprietary hashing algorithm produced by Thorn.**
- **PhotoDNA and internal proprietary tools: a combination of technology solutions are used to surface accounts violating our rules on Child Sexual Exploitation.**
- **Media Risk Scanning: We receive a media classifier score through Safer and it is used to filter false positive hash matches at the moment. We use a novel classifier model to rate media shared through posts' likelihood of being CSAM. Media that receives a high score is then sent to human review.**

  In addition, X is currently beta testing a text-based machine learning classifier developed by Thorn to assist in the detection of sextortion, CSAM, child-access, and child sexual abuse discussion.

b. What benefits can AI provide to helping detect and/or stop harmful content to children online?

AI can assist in proactive detection, hash-matching, analysis of

media, analysis of text, and automation of enforcement, to name a few benefits.

c. What does a human reviewer provide that an AI or automated tool cannot? Will we always need some measure of human review in assessing online harms to children?

Human capacity is essential for understanding the context of cybercrimes, investigating account behaviors, analyzing networks of bad actors, providing qualitative feedback on processes, and working with cross-functional teams to enforce policies.

d. The FTC has sent mixed signals in its enforcement of COPPA. While the Commission emphasizes not over relying on use of automated tools or AI, they have nonetheless found liability for using human review as alternative signaling overreliance on automated tools. What improvements, if any, should Congress make to clarify the legal tension between use of automated detection tools vs. human review?

X has billions of accounts on its platform; robust, automated tools are essential to implement its safety policies—including policies related to COPPA—at scale. We encourage regulators and lawmakers to create the space for platforms such as X to innovate and improve these automated tools, including by protecting good faith efforts to improve online safety from legal challenges.

5. In 2021, Congress directed the FTC to research and report on how AI can be used positively to detect and combat fraudulent or deceptive content online. Rather than viewing AI as a potential solution to our online woes, the FTC instead issued a report that read more like an indictment of the technology.

   a. Please explain whether, in your view, AI can be used to positively detect and combat fraudulent or deceptive content, including the recent use of deepfakes or other scams to harm consumers.

   Yes.

   b. Has the FTC ever consulted with your company to learn how your company deploys AI to better detect and combat fraudulent or deceptive content? Has the DOJ? How about the Federal Elections Commission?

To the best of our knowledge, X has not been consulted on these issues.

c. How can Congress empower agencies to use AI positively for the protection of American consumers from fraudulent or deceptive content?

AI is an important and developing technology for detecting and combatting fraudulent and deceptive content. Both the private sector and government agencies should carefully assess how they can deploy that evolving technology most effectively. Congress can help to ensure that federal agencies have the personnel and resources necessary to evaluate and deploy that technology most effectively.

6. Please provide a description of your company's policy regarding the sale or transfer of the data of American users collected on your platform to a third party, including data brokers.

As stipulated in Section 6 of our Privacy Policy (https://twitter.com/en/privacy), we have specific legal bases for collecting, using and sharing user data (further described here https://help.twitter.com/en/rules-and-policies/data-processing-legal-bases) but do not sell user's personal data.

7. Has your company ever sold the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the categories of data sold.

In response to your inquiry, our teams are investigating this question and will be happy to follow up with you.

8. Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the circumstances underlying the basis for such transfer.

Not to our knowledge, with the exception of an incident several years ago under prior management in which malicious actors inside the company transferred data concerning a discrete number of users to the government of Saudi Arabia without the company's knowledge or consent. The company cooperated with U.S. authorities in connection with the incident.

9. Has your company ever sold the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and the categories of data sold.

In response to your inquiry, our teams are investigating this question and will be happy to follow up with you.

10. Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and categories of data transferred.

Not to our knowledge.

11. Does your company have a policy to restrict third party use and/or transfer of data collected from users on your platform? Please be specific, including how you enforce such restrictions and whether such restrictions prohibit the sale or transfer of such data to a government agency, including a foreign government agency.

X only permits third-party access to aggregate data via our approved API process. When a developer seeks access to X's API, they are required to abide by the terms of our developer agreement upon sign up. (ttps://developer.twitter.com/en/developer-terms/agreement-and-policy), a legally binding agreement between the developer and X. X specifically prohibits the use of user data by any entity for surveillance purposes, investigating or tracking X users or their content or in any other way inconsistent with users' privacy expectations as described in our Privacy Policy (https://twitter.com/en/privacy). Should we learn that a developer has violated these policies, we will take appropriate action, which may include suspension and termination of access to X's APIs, reporting to relevant authorities, and/or legal action as appropriate. Further information on restricted uses of X's API can be found at https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases.

12. Between July 4, 2023 and July 14, 2023, was your company contacted by any employee of or contractor for any of the following agencies? Please answer "yes" or "no" for each agency and, if "yes," provide the date(s) of contact and the name(s) of the agency employees or contractors that contacted your company.

X's responses to questions 12(a)-(l) are based on a reasonable search of its systems and exclude routine requests for support with X accounts.

a. U.S. Department of Health and Human Services (HHS)

X is not aware of any outreach by an employee or contractor of HHS during the specified time period.

b. National Institute of Allergy and Infectious Diseases (NIAID)

X is not aware of any outreach by an employee or contractor of NIAID during the specified time period.

c. Centers for Disease Control and Prevention (CDC)

X is not aware of any outreach by an employee or contractor of CDC during the specified time period.

d. U.S. Food and Drug Administration (FDA)

X is not aware of any outreach by an employee or contractor of FDA during the specified time period.

e. The National Institutes of Health (NIH)

X is not aware of any outreach by an employee or contractor of NIH during the specified time period.

f. U.S. Department of Homeland Security (DHS)

X is not aware of any outreach by an employee or contractor of the DHS during the specified time period.

g. DHS Cybersecurity and Infrastructure Security Agency (CISA)

Certain X employees received mass emails sent to CISA listservs, including notifications of trainings and public cybersecurity vulnerability notifications. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of CISA during the specified time period

h. U.S. Census Bureau

X is not aware of any outreach by an employee or contractor of the Census Bureau during the specified time period.

i. Federal Bureau of Investigation (FBI)

X engaged in communications with the FBI related to routine law enforcement requests, and received notifications of webinars and other industry-wide training opportunities, during the specified time period. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of the FBI during the specified time period.

j. U.S. Department of Justice (DOJ)

X engaged in communications with the DOJ related to routine law enforcement requests during the specified time period. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of the DOJ during the specified time period.

k. The White House Executive Office of the President (EOP)

X is not aware of any outreach by an employee or contractor of the EOP during the specified time period.

l. U.S. Department of State

Certain X employees received emails from State Department listservs during the specified time period. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of the Department of State during the specified time period

13. Is it your company's policy to prevent children under 13 from using your social media app(s) or creating an account?

Yes.

14. In your view, would it be appropriate for school-aged children to spend time on or access your company's social media app(s) during class?

Technology usage in the classroom is a matter of concern for the school district, school administration, the teacher, and parents.

15. As a parent, would you be concerned if your child were able to access your company's social media app(s) during class via a school network or device?

Technology usage in schools should be managed by school administration, teachers, and parents.

16. In your view, should elementary and secondary schools block students' access to your company's social media app(s) on school networks and devices?

We defer to school administrators on which technologies to allow and which apps to block on their respective networks and devices.

17. Do you think that school buses equipped with Wi-Fi should allow children to access your company's social media app(s) via a school bus Wi-Fi network during their rides to and from school?

We defer to school administrators on which technologies to allow and which apps to block on their respective networks and devices.

18. As a parent, do you think it is important to supervise your children's internet access?

Yes.

19. As a parent, would you be concerned if your child's school allowed your child to access the internet on an unsupervised basis, such as on your child's bus ride to and from school via the school bus Wi-Fi?

Yes.

20. Do you think Congress should require schools, as a condition of receiving broadband subsidies through the Federal Communications Commission's E-Rate program (which funds broadband for elementary and secondary schools), to block students' access to your company's social media app(s) from school-run networks?

We generally do not oppose efforts to regulate access to social media platforms in schools or on school-run networks.

21. Do you support the bipartisan *Eyes on the Board Act of 2023*, S. 3074?

While we have not taken an official position on this legislation, we generally have not opposed legislative efforts regulating access to social media platforms in schools.

22. Have you, your company, or any foundation associated with you or your company, donated or contributed funding, equipment, or services to any of the following organizations in the last ten years (CY 2013 to CY 2023)?

    a. Education and Libraries Networks Coalition (EdLiNC) - **No**

    b. Open Technology Institute - **No**

    c. Consortium for School Networking (COSN) - **No**

    d. Funds For Learning - **No**

    e. State Educational Technology Directors Association (SETDA) - **No**

    f. Schools, Health, and Libraries Broadband Coalition (SHLB) - **No**

    g. State E-Rate Coordinators' Alliance (SECA) - **No**

    h. EducationSuperHighway - **No**

    i. All4Ed - **No**

    j. Public Knowledge - **Yes**

    k. Fight for the Future - **No**

    l. Free Press - **No**

    m. Electronic Frontier Foundation - **No**

    n. Benton Foundation or Benton Institute for Broadband & Society - **No**

    o. Electronic Privacy Information Center - **No**

23. For each such donation or contribution described in the prior question, please detail (1) the type of donation or contribution, such as financial donation, goods or equipment, services, etc.; (2) who made the donation or contribution; (3) the recipient organization; (4) the year the donation or contribution was made; and (5) the total value of that donation or contribution.

Public Knowledge - $10,000 in 2021 donated from Twitter

24. Yes or no: Did employees of or contractors for the Cybersecurity and Infrastructure Security Agency (CISA) ever ask Twitter/X to meet with employees of or contractors for the Department of Homeland Security Office of Inspector General (DHS OIG)?

To the best of my knowledge, X has not received any such request from CISA.

a. If yes, provide the date of the request from CISA, the channel through which the request was made, and the name of the CISA employee(s) or contractor(s) who made the request.

Questions from Senator Thom Tillis
for the CEO of X Corp., Ms. Linda Yaccarino

1. Twenty-one is the minimum age to purchase highly regulated adult products such as alcohol, tobacco, and nicotine. Nevertheless, there is a proliferation of user-generated content posted on social media sites featuring underage use of these products.

   Recently, some have proposed banning these age-restricted products due in part to the user-generated content being available on your respective platforms. Surely, banning these products cannot be the answer. However, we must do more – your company must do more – to shield underage audiences from exposure to this content.

   Therefore, as the content moderator of these platforms, what policies do you have in place, and what more can you do, to prevent this type of user-generated content from reaching underage audiences? How do you respond to requests to pull this content from your sites when deemed inappropriate for underage audiences?

   X prohibits knowingly marketing or advertising the following products and services to minors:

   - **Alcoholic beverages and related accessories**
   - **Weapons, ammunition, or weapons training/certification**
   - **Projectile, BB, or pellet guns/devices***
   - **Fireworks***
   - **Aerosol paint or etching cream capable of defacing property**
   - **Tobacco products or accessories, including electronic cigarettes***
   - **Any controlled substance or paraphernalia***
   - **Drug paraphernalia***
   - **Any substance or material containing Salvia divinorum or Salvinorin A***
   - **Weight loss products and services and content focused on weight loss**
   - **Health and wellness supplements (including, but not limited to, health, dietary, food, nutrition, weight loss, and muscle enhancement substances and supplements)**
   - **Tanning in an ultraviolet tanning device**
   - **Gambling products and services, including lotteries**
   - **Body branding such as tattooing, body piercing, or permanent cosmetics**
   - **Sexual products and services, or content that is adult in nature**
   - **Please note that asterisked items are prohibited from advertising on X overall.**

   While posts promoted through X's advertising services are labeled as "Promoted" and must abide by our X Ads Policies, organic, non-promoted posts may also be considered paid product placements, endorsements, or advertisements ("Paid Partnerships").

The following are examples of Paid Partnerships:

- **A user, including a creator or brand, has been or may be compensated for a post (including in the form of money, gifts, loans of products, or other rewards or incentives), or**
- **A post is created as part of, or in connection with, a commercial relationship (such as a current or recent 'brand ambassador' arrangement), or**
- **A post includes an affiliate link or discount code through which the user might receive some kind of benefit, incentive or reward**

Posts that are part of a Paid Partnership posted as an organic post will require clear and prominent disclosures indicating the commercial nature of such content. For example, "#ad", "#paidpartnership", "#sponsored".

Failure to include an appropriate disclosure in a clear and prominent way could result in enforcement actions. In addition to abiding by the X Rules, users, including creators and brands, that participate in Paid Partnerships are responsible for complying with all applicable laws and regulations, including but not limited to, all advertising laws and, where applicable, FTC regulations including the FTC's Guides Concerning the Use of Endorsements and Testimonials in Advertising.

2. Public reports conclude that drug cartels use social media like TikTok, META, X, Snapchat, and others to plan, organize, and communicate in real-time. These communications coincide directly with criminal activity.

   What are your companies doing to crack down on cartel coordination? Specifically, in the recruitment of children to commit crimes or assist in the sale/distribution of illicit drugs?

   We are aware of general reports of cartels utilizing certain social media platforms to advertise to and recruit minors, however, we are unaware of any such coordinated campaigns on X. We remain vigilant of this activity and if your office or any stakeholder has evidence of this behavior on X, please share immediately with our teams so that we may investigate.

   X maintains a robust cybercrime policy on Illegal and Regulated Goods (IRGS) and Services. Under this policy, users may not use our service for any unlawful purpose or in furtherance of illegal activities. This includes selling, buying, or facilitating transactions in illegal goods or services, as well as certain types of regulated goods or services.

   In addition to reports received, we proactively surface activity that may violate this policy for human review.

Goods or services covered under this policy include, but are not limited to:

- **counterfeit goods and services;**
- **drugs and controlled substances;**
- **human trafficking;**
- **products made from endangered or protected species;**
- **sexual services;**
- **illicitly obtained materials; and**
- **weapons, including firearms, ammunition, and explosives, and instructions on making weapons (e.g. bombs, 3D printed guns, etc.**

If we determine that a user violated this policy, we may suspend their account, including upon first review.

Accounts that appear to be using misleading account information in order to engage in spamming, abusive, or disruptive behavior to promote the sale of illegal and regulated goods and/or services may be subject to suspension under our platform manipulation and spam policy.

3. What steps does your platform take to proactively remove, delist, and ban any posts, users, websites, and advertisements associated with the sale and distribution of fentanyl and other illicit drugs?

Under our IRGS policy outlined above, we use a mix of proactive detection and user reporting to enforce potential violations. We proactively detect content for human review using a mix of technology and heuristics depending on the activity. We utilize databases of known terms and slang, drug names, and emojis, for example, to inform our detection models. If a particular link is found to be harmful, we denylist that URL and block it from being posted at all. More information on our approach to harmful links can be found here: https://help.twitter.com/en/safety-and-security/phishing-spam-and-malware-links

4. One area of growing concern is the sale and distribution of fake or counterfeit vaping devices online, particularly in connection with so-called Delta-8 THC. Counterfeit vapes, many coming from China, have unsafe and even potentially deadly chemicals. They have caused hospitalizations and death. What are your platforms doing to combat this problem?

See above answer to Question 1 regarding prohibited advertising to minors, which captures tobacco accessories, including e-cigarettes. Such activity could also be captured by our Illegal and Regulated Goods Policy, as outlined in our answer to Question 2 above.

5. What are the main impediments your platform encounters in identifying all fentanyl and illicit drug advertisements posted to your platform(s) automatically? Please describe any circumstances in which you do not or cannot apply detection

technologies against content transmitted on your platform(s).

We are constantly soliciting input from experts and law enforcement on the latest trends, nomenclature, and slang involving the trafficking of illegal goods and services. The information we receive from these entities feeds into our proactive detection capabilities.

6. How many posts, users, websites, and advertisements have you removed, delisted, and banned per year for the sale and distribution of fentanyl and other illicit drugs? How many per year? Have you seen an increase in illicit drugs being advertised to children on your platform(s)?

In 2023, we suspended approximately 630,000 accounts and removed more than 1.8 million posts under our Illegal and Regulated Goods Policy.

7. Are there any other roadblocks or impediments that you face in addressing fentanyl and illicit drug advertisements on your platform(s), and working with law enforcement on such matters? If yes, what are they? If no, how many cases have been transmitted to law enforcement and DEA?

We are constantly soliciting input from experts and law enforcement on the latest trends, nomenclature, and slang involving the trafficking of illegal goods and services. The information we receive from these entities feeds into our proactive detection capabilities.

In response to your inquiry regarding transmissions to law enforcement, our teams are investigating this question and will be happy to follow up with you.

8. How do you work with organizations, advocates, and experts focused on drug prevention and addiction recovery to adapt your products and operations to keep up with the illicit drug crisis — including working with parents that have lost children due to lethal drugs bought online?

We work with organizations around the world dedicated to supporting recovery and online safety. One notable partner in the US in this work is Mobilize Recovery, and we are honored to support their campaigns on X via advertising credits.

9. What are the total number of meetings that your company has had with parents to address online safety concerns? Can you provide the total number of meetings over the last three years? Please separate this last question's answer by number per year.

We have had numerous meetings around the globe with groups dedicated to the safety of minors online, including parents. We do not have a specific number of meetings we are able to share. We are committed to continuing to meet with

advocates for child protection and gather feedback on how we can make X safer for minors.

10. In 2022, then National Center for Missing & Exploited Children (NCMEC) received over 32 million reports of Child Sexual Abuse Material (CSAM). Reports of online sex crimes to the CyberTipline are growing exponentially year by year. Out of those 32 million reports, how many did your platform submit to NCMEC?

In 2022, Twitter submitted approximately 98,000 reports to the CyberTipline. In 2023, X submitted approximately 850,000 reports to the CyberTipline.

In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year.

Since April of 2023, we have increased training for content moderators on the tools and policies for NCMEC reporting. In turn, this has led to a 10x increase in the volume of manually-submitted NCMEC reports, from an average of 6,300 reports per month to an average of 64,000 reports per month from June through November 2023. We are evaluating more sources of potential CSAM than we could before.

11. There is concern that this number is going to fall dramatically this year because of the adoption of end-to-end encryption, not because the problem is going away. How will your company track and address this issue moving forward?

At this time, X does not offer full end-to-end encryption. Encrypted Direct Messages are limited to only subscribers of X Premium, and only text is encrypted. Encrypted conversations will appear as separate conversations, alongside your existing Direct Messages in your inbox. Direct Messages are not defaulted to encrypted.

Users need to satisfy the following conditions in order to send and receive encrypted messages:

i. **both sender and recipient are on the latest X apps (iOS, Android, Web);**
ii. **both sender and recipient are verified users or affiliates to a verified organization; and**
iii. **the recipient follows the sender, or has sent a message to sender previously, or has accepted a Direct Message request from the sender before.**

12. Has your platform seen an increase of suspected online child sexual exploitation-CSAM over the past few years? If so, what do you believe is the driving factor on why it's happening on your platform?

Generally, we have not seen a marked increase in the prevalence of CSAM, however, we have improved our capabilities of detecting CSAM, made it easier to report CSAM across all product surfaces, increased training of agents, implemented more automated technologies, and become more aggressive in automated enforcement of accounts that engage with the content.

13. What are some new tools or strategies that your platform has implemented to identify CSAM? How closely does your platform work with NCMEC?

We are constantly seeking feedback and input from trusted organizations that are aligned in the mission to combat online CSE. Foundational to our work is our multidimensional partnership with NCMEC, which manages the CyberTipline program, regularly convenes global stakeholders and facilitates actionable feedback from law enforcement that makes us better. We have quarterly operational and policy syncs and members of our teams are connecting every month to share information and feedback on our reporting. Other instrumental partners are the Tech Coalition and WeProtect, alliances that push our innovation and provide critical information sharing on emerging threats and behaviors.

In December 2022, we launched a new product partnership that allows us to take down more violative media than before. Built by Thorn, Safer allows tech platforms to identify, remove, and report child sexual abuse material at scale.

We are investing in products and people to bolster our ability to detect and action more content and accounts, and are actively evaluating advanced technologies from third-party developers that can enhance our capabilities. Some highlights include:

a. **Automated NCMEC reporting: In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year.**
b. **Expanded Hash Matching to Videos and GIFs: For the first time ever, we are evaluating all videos and GIFs posted on X for CSAM. Since launching this new approach in July 2023, we have matched over 70,000 pieces of media.**
c. **Launched Search Intervention for CSE Keywords: CSAM impressions occur more on search than on any other product surface. In December 2022, we launched the ability to entirely block search results for certain terms. We have since added more than 2,500 CSE keywords and phrases to this list to prevent users from searching for common CSE terms.**

14. What resources or help does your platform provide to victims of CSAM? Does your platform work with local victim groups and professionals?

The uniqueness of X is the role it serves as a platform for public conversation, the global town square of the internet. X has always been a place for victims to bring awareness to their causes and issues of public concern, like legislation. We will continue to support organizations around the world that promote online safety and we welcome any recommendations of victims groups that we could support in their campaigns and advocacy.

15. What are the top technical hurdles your company faces in combating CSAM?

Cross-platform bad actors always pose a challenge as they may attempt to direct people to other platforms where CSAM is exchanged or transactions happen. Sharing signals between platforms will assist in rooting out these types of behaviors, and that is why we have applied to the Tech Coalition's Project Lantern.

16. There seem to be competing views on how to regulate algorithms. Some suggest that more transparency is needed, while others want more privacy. Can you provide your perspective on whether more or less transparency is needed when it comes to algorithms?

In 2023, we published our recommendation algorithm and a comprehensive blog explaining our recommendation system.
https://blog.twitter.com/en_us/topics/company/2023/a-new-era-of-transparency-for-twitter

You can find the code to our recommendation algorithm on GitHub.
https://github.com/twitter/the-algorithm

17. Do you believe that large companies and platforms like yours can use algorithms to stifle innovation or small businesses?

We have published our recommendation algorithm in an effort to be transparent about how our systems work and to encourage innovation.

18. What do you believe is the role of government in regulating algorithms? What, if any, unintended consequences would there be if Congress gets involved?

Foundational to regulation of algorithms, or artificial intelligence, is strong privacy protections for individuals. Congress should begin by passing comprehensive privacy reform. We also believe it is important to give people control, which is why we offer all users the 'Following' experience, as an alternative to the algorithmic 'For You' tab.

19. Are you aware of your platform using surveillance advertisements to target children (anyone under the age of 18) with specific ads? If so, in your opinion, how can this be mitigated?

To the best of my knowledge, X does not use surveillance advertisements to target children with specific ads. Generally, you may not specifically target advertisements to 13-17 year olds on X.

20. Beyond surveillance advertisements, are there any other algorithmic-based practices being implemented that are particularly detrimental to children? In your opinion, how can this be mitigated?

To the best of my knowledge, I am not familiar with algorithmic-based practices being implemented that are particularly detrimental to children.

21. Are you aware of any surveillance advertisements or algorithms that are used to target children, specifically to promote drugs and the sale of narcotics?

To the best of my knowledge, I am not aware of surveillance advertisements or algorithms that are used to target children on X.



1 Hacker Way
Menlo Park, CA 94025
United States

April 19, 2024

Chairman Richard Durbin
Ranking Member Lindsay Graham
US Senate Committee on the Judiciary
224 Dirksen Senate Office Building
Washington DC, 20510

Dear Chairman Durbin, Ranking Member Graham, and Members of the Committee:

Enclosed is Meta's complete submission of responses to the questions for the record from the Senate Judiciary Hearing entitled "Big Tech and the Online Child Sexual Exploitation Crisis" held on January 31, 2024.

Sincerely,

Meta Platforms, Inc.

Questions from Senator Durbin

Question 1. **For each year from 2019 to 2023 for Meta and all subsidiaries, please provide the following:**

a. the total number of users on your platforms;

In annual filings, we report estimates of the numbers of daily active people based on the activity of people who visited at least one of Facebook, Instagram, Messenger, and WhatsApp—there were an average of 3.19 billion active people in December 2023. In these metrics, we do not seek to count the total number of accounts across our services because we believe that would not reflect the actual size of our community.

b. the total number of users under the age of 18 on your platforms;

When we look at self-reported ages of our US daily active users, about 6% of Instagram accounts belong to teens under 18, and 1% of Facebook accounts belong to teens under 18.

Meta recognizes the need to keep people who are too young off Facebook and Instagram. Both Facebook's Terms of Service and Instagram's Terms of Use in the United States require people to be at least 13 years old to sign up for Facebook and Instagram. Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

Anyone, including individuals who do not have a Facebook or Instagram account, can report to Meta that someone is or appears to be under the age of 13 by filling out a platform-specific online reporting form. Reported accounts are evaluated either by an automated process or through human review, and in some cases both. Our content reviewers are trained to confirm and remove accounts that appear to be used by people who are underage. While manual review is a labor- and time-intensive process, Meta has taken steps to review accounts flagged as potentially underage as quickly as possible after they are reported. Additionally, Meta may learn that someone is underage directly from the individual, if the person attempts to change the date of birth on their account to a date that would make them under 13. In this case, the individual is automatically placed in an "age checkpoint," and we remove the account if the person cannot verify they are over 13.

Identifying Potentially Underage Accounts

Anyone, including individuals who do not have a Facebook or Instagram account, can report to Meta that someone is or appears to be under the age of 13 by filling out a platform-specific online reporting form. Reported accounts are evaluated either by an automated process or through human review, and in some cases both. Our content reviewers are trained to confirm and remove accounts that appear to be used by people who are underage. While manual review is a labor- and time-intensive process, Meta has taken steps to review accounts flagged as potentially underage as quickly as possible after they are reported. Additionally, Meta may learn that someone is underage directly from the individual, if the person attempts to change the date of birth on their account to a date that would make them under 13. In this case, the individual is automatically placed in an "age checkpoint," and we remove the account if the person cannot verify they are over 13.

Automated Evaluation

An account that has been flagged as potentially underage will first go through an automated process that determines whether the account should be escalated for human review or immediately allowed to continue using the platforms. Where Meta has evidence indicating that the reported individual is over the age of 13, Meta may automatically permit the person who has been flagged as potentially underage to continue using Facebook or Instagram. For example, this can occur when a human reviewer previously evaluated the account for potential underage usage and approved the individual to continue using the platform following the review (pursuant to review guidelines detailed below), the account was previously placed in an age checkpoint and the person submitted sufficient documentation demonstrating they were at least 13 years old, or the account is so old that it could not reasonably belong to a person under 13. A flagged account will also be permitted to remain on the platform if the account contains no bio or photos, because, as discussed below, a reviewer relies on this data to evaluate whether the account belongs to an underage person.

Manual Review of Potentially Underage Accounts

Flagged accounts that cannot be resolved through the automated processes described above are directed to human reviewers for further evaluation. Meta employs tens of thousands of human reviewers whose duties include reviewing these Facebook and Instagram accounts to manually look for signs that an account has violated the applicable terms of service or content guidelines, including accounts suspected as belonging to people under 13.

All potentially underage accounts that are manually reviewed are evaluated to determine whether they meet our age requirements. For example, reviewers consider the following factors:

- **Account Bio**: Reviewers first evaluate the account's bio for contextual information or self-admission about a person's real age, including a written statement of the person's age, birth date, or grade in school. Reviewers are also trained to look for additional signals that indicate whether the account holder is underage. An account that contains information that explicitly states or contextually implies that the individual is under 13 will be checkpointed and the person will be required to provide Meta with proof of age.
- **Account Photos**: If the account bio does not contain sufficient written information to determine age, the reviewer will evaluate the photos contained in the account. If a human reviewer finds sufficient signals that the account holder may be under 13, or is unsure of whether an account holder is under 13 based on a review of the account media, the account will immediately be placed in an age checkpoint.

Responding to Potentially Underage Accounts

When Meta identifies a potentially underage account, their account will be placed in an age checkpoint. While in the checkpoint, a person does not have access to their account, and they are shown a blocking screen if they attempt to log into their account. This means checkpointed accounts cannot view or interact with any content or ads on the platform. Additionally, checkpointed accounts are not visible to other people on the platform, and people cannot see or interact with the checkpointed account or the photos or videos posted to it.

If the person is unable to demonstrate that they are 13 years of age or older, their account is permanently disabled and removed, and the data is deleted consistent with Meta's standard deletion policies.

Other Mechanisms for Identifying Potentially Underage Accounts

In addition, we have also partnered with Yoti, a company that offers privacy-preserving ways to verify age. Yoti is verified by the Age Check Certification Scheme and is the leading age verification provider for several industries around the world including social media, gaming and age-restricted e-commerce. Expert and governmental organizations in youth and privacy have publicly endorsed Yoti for their approach and expertise in responsible artificial intelligence.

For example, on both Instagram and Facebook, a person who attempts to change their date of birth to go from the age of under 18 to 18 or older is required to verify their age through one of two options,[1] ID upload or video selfie provided by the third-party Yoti. If Yoti estimates that the person is under the age of 13, the account will be placed in an age checkpoint. As explained above, if the person is unable to demonstrate that they are 13 years of age or older, their account

[1] And on Instagram, a person can ask mutual friends to verify their age.

is permanently disabled and removed, and when the account is disabled, the data is deleted consistent with Meta's standard deletion policies.

c. the estimated number of users under the age of 13 on your platforms;

In the last two quarters of 2021, Meta removed more than 4.8 million accounts on Facebook and 1.7 million accounts on Instagram because they were unable to meet our minimum age requirement. For more information on our industry-first work to find and remove accounts of people under 13 from Facebook and Instagram, please see the response to your Question 1(b).

d. the number of users who are using any of the parental supervision tools offered through your Family Center?

We built a Family Center to help teens and families build healthy online habits. The Family Center is a central place for parents and guardians to access supervision tools and resources from leading experts. It includes an education hub where parents and guardians can access resources from experts and review articles, videos, and tips on topics like how to talk to teens about social media. Parents and guardians can also watch video tutorials on how to use these new supervision tools. Our vision for the Family Center is to eventually allow parents and guardians to help their teens manage experiences across Meta technologies, all from one place.

Among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians and teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

Nonetheless, it can be challenging for parents to supervise the many apps that their teens may use, which is one of the reasons we support federal legislation at the app store level that would make it simpler for parents to oversee their teens' online lives.

e. your company's annual revenue;

In 2023, Meta's annual revenue was $134 billion.

f. your company's annual budget for trust and safety;

Our budget for safety and security is now greater than the whole revenue of our company at the time that we went public in 2012. Meta has invested more than $20 billion in safety and security across our platforms since 2016, and $5 billion in 2023 alone. In 2022, Meta invested approximately $6 billion. In 2021, Meta invested about $5 billion.

g. your company's annual budget to address online child sexual exploitation;

We do not segment out our safety budget in this way. Our investment in this space often overlaps across harm types, and has allowed us to build tools and technology that are used to combat a range of different types of abuse. For example, we invest heavily in fighting fraud on our platform, including scams like financial sextortion, a form of online sexual exploitation. We are unable to provide a precise estimate of different child safety related budgets, as this work is embedded throughout the company.

That said, we remain focused on advancing our industry-leading integrity efforts and continue to invest in teams and technologies to protect our community. We are committed to continuing our work to protect teens, obstruct criminals, and support law enforcement in bringing them to justice. Meta spent around $5 billion on safety and security in 2023.

For more information about our investments in safety and security, please see the response to your Question 1(f).

h. the total number of employees working to address trust and safety;

Since 2016, Meta has significantly expanded the number of people who work on safety and security. By 2018, Meta doubled the number of people who work on safety issues from 10,000 to 20,000, which includes content reviewers, systems engineers, and security experts. By 2020, Meta built a global team of 35,000 people to work on safety and security, including specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to review potentially violating content and report findings to NCMEC. And by 2022, Meta had more than quadrupled the number of people working on safety and security since 2016 to over 40,000 people. We continue to have around 40,000 people devoted to safety and security efforts.

Reductions in workforce occurred across the entire tech industry, but to be clear, the restructuring efforts of last year *do not* change the commitment we have to our ongoing child safety efforts at Meta. Providing teens and their families with safe experiences is one of our most important priorities. We remain focused on advancing our industry-leading integrity efforts and continue to invest in teams and technologies to protect our community. We are committed to continuing our work to protect teens, obstruct criminals, and support law enforcement in bringing them to justice."

i. the total number of employees working to address online child sexual exploitation.

We do not segment out our workforce numbers in this way. Our investment in this space often overlaps across harm types, and has allowed us to build tools and technology that are used to combat a range of different types of abuse. For example, we invest heavily in fighting fraud on our platform, including scams like financial sextortion, a form of online sexual exploitation. That said, we continue to have around 40,000 people devoted to safety and security efforts. For more information about the number of employees working on safety and security, please see the response to your Question 1(h).

Question 2. **How did your company determine that 13 was the appropriate age for a child to begin using your platform?**

The minimum age requirement of 13 is relatively standard across our industry. We develop our policies and services not only to comply with COPPA but also to meet and exceed the high standards of parents and families. Our policies are developed by our policy team in close consultation with our safety teams, compliance teams, community operations teams, among others. We also consult with external stakeholders and experts in fields like child safety, privacy, technology, public safety, and more.

Question 3. **What legal obligation does your company have in the United States to ensure that your platforms are safe for children before they are launched?**

We develop our policies and services not only to comply with COPPA but also to meet and exceed the high standards of parents and families. Our policies are developed by our policy team in close consultation with our safety teams, compliance teams, community operations teams, and more—and we consult with external stakeholders and experts in fields like child safety, privacy, technology, public safety, and more.

We go beyond legal requirements and use sophisticated technology to proactively seek out abusive material, and as a result, we find and report more inappropriate content than anyone else in the industry. As the National Center for Missing and Exploited Children (NCMEC) put it, Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs."

At Meta, we are proud to work closely with trusted organizations and individuals in an effort to help support families in fostering positive online relationships. We have worked closely with leading child development experts, educators, and parents. We created an advisory board of experts. With them, we have considered important questions like: Is there a "right age" to introduce kids to the digital world? Is technology good for kids, or is it having adverse effects on

their social skills and health? Our team of advisors includes top experts in the fields of child development, online safety, and children's media currently and formerly from organizations such as the Yale Center for Emotional Intelligence, Connect Safely, Center on Media and Child Health, Sesame Workshop, and more. These advisors are helping us grow our knowledge and guide us as we develop services.

We also support federal legislation that requires app stores to get parents' approval whenever their teens under 16 download apps. With this solution, when a teen wants to download an app—including ours—app stores would be required to notify their parents, much like when parents are notified if their teen attempts to make a purchase. Parents can decide if they want to approve the download. Parents want this type of clear system for parental control over what apps their kids are using. For example, 3 out of 4 parents favor introducing app store age verification, and 4 out of 5 parents want legislation that we support, requiring app stores to get parental approval whenever teens download apps.

***Question 4*. For users under the age of 18,**

- **a. what are the default privacy settings for their accounts?**
- **b. what limitations are placed by default on content these users can access, content that will be directed toward them, and individuals they can communicate with?**
- **c. can they change their default settings without the awareness of their parent or guardian, or without the consent of their parent or guardian?**
- **d. in 2023, how many changed their default settings?**

With respect to default privacy settings, in the US, everyone who is under 16 years old is defaulted into a private account when they join either Instagram or Facebook. On Instagram, if a person has a private account, people have to request to follow them to see their posts, Stories, and Reels unless they choose to allow others to reshare their content. People also cannot comment on their content in those places, and they will not see their content at all in places like Explore or hashtags. We also have other defaults in place when people under 18 first sign up for Instagram, including not allowing people they do not follow to tag or mention them, or include their content in Reels Remixes or Guides. And for Facebook, everyone who is under the age of 16 in the US is defaulted into more private settings when they join Facebook, including restricting:

- Who can see their friends list;
- Who can see the people, Pages and lists they follow;
- Who can see posts they are tagged in on their profile;
- Who is allowed to comment on their public posts; and

- Minors' contact info, school, and birthday from appearing in search to a public audience.

Additionally, to help protect teens from unwanted contact, in the US, we have turned off the ability for teens under the age of 16 to receive direct messages from anyone they do not follow or who they are not connected to on Instagram—including other teens—by default. Under this default setting, teens can only be messaged or added to group chats by people they already follow or are connected to, helping teens and their parents feel even more confident that they will not hear from people they do not know in their direct messages. Teens under 16 in supervised accounts will need to get their parent's permission to change this setting. We are also making these changes to teens' default settings on Messenger, where in the US, people under 16 will only receive messages from Facebook friends, or people they are connected to through phone contacts, for example.

In addition to these default privacy settings, on the content front, we restrict the visibility of certain types of age-inappropriate content on Facebook and Instagram for people under the age of 18, including content related to alcohol, tobacco, bladed weapons, weight loss products, cosmetic procedures, sex toys, sexual enhancement products, gambling and entheogens. We also restrict the visibility of content depicting a person who engaged in euthansia/assisted suicide in a medical setting to only adults over the age of 18, and include a sensitivity screen. And we are working to remove more age-inappropriate content (such as someone posting about their ongoing struggle with thoughts of self-harm) from teens' experiences on Instagram and Facebook, even if it is shared in Feed and Stories by someone they follow. We already aim not to recommend this type of content to teens in places like Reels and Explore. When people search for terms related to suicide, self-harm and eating disorders, we hide these related results and direct them to expert resources for help.

We also automatically place teens into the most restrictive content control setting on Instagram and Facebook. We apply this setting for teens when they join Instagram and Facebook, and are now expanding it to teens who are already using these apps. Our content recommendation controls—known as "Sensitive Content Control" on Instagram and "Reduce" on Facebook—make it more difficult for people to come across potentially sensitive content or accounts in places like Search and Explore. Parents of teens under 16 using supervision tools will be prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

A large majority of teens keep their default settings. For example, among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians and teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

Question 5. **If the default settings are different for users aged 16 and 17 than they are for users under the age of 16, please explain why your company takes this position, how this position was developed, and whether any company personnel voiced objections to or raised concerns about this position.**

Research and expert consultation play a major role in Meta's product development process, including in helping to ensure that certain designs and tools are age-appropriate. Youth are not a monolithic group—they vary in age, maturity, home situations, cultural norms, and parental support. We regularly consult with experts in fields such as child development, mental health, and digital literacy to help understand how to apply age-appropriate defaults for teens, in line with their developmental needs.

Some of our defaults are set at the age of 18. For example, we restrict all teens under the age of 18 from accessing the "More" setting of our Sensitive Content Control and we default teens under the age of 16 into the most restrictive content and recommendations settings. 99% of teens who are defaulted globally and in the US are still using this setting a year later.

Question 6. **An article published in the Washington Post the day before the hearing indicates that by the end of 2022, less than 10 percent of teens on Instagram had enabled the parental supervision setting. Of those who did, only a single-digit percentage of parents had adjusted their kids' settings. A Meta spokesperson is quoted as saying "we're always working to make sure parents know about and can choose to use these features."**

a. **What studies, research, summaries, or data does your company, including subsidiaries, have reflecting the efficacy of its parental controls and child safety measures? Please provide these studies, research, summaries, or data.**
b. **Please describe you work to make sure parents know about and choose parental supervision tools.**

We have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being

granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. For example, Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

We are constantly working to make sure parents know about and can choose to use parental control features. We reach parents in a variety of ways, including through Family Center's Education Hub, advertising campaigns, in-app promotion, our work with safety partners, and events with parents. We also collect input from teens in multiple ways. For example, we have hosted programs with organizations such as Girls Scouts of America and the National Parent Teacher Association to create awareness of our safety tools and to get feedback from teens. We have also launched TTC Labs, a global co-design program, that invites young people, parents, and experts to participate as collaborators in our design process, empowering them to have their say and ensuring our products meet their needs.

Our ads encouraging parents to use our youth well-being tools and features were seen more than one billion times by tens of million people in the United States since August 2022. Additionally, in 2023, we hosted "Screen Smart" events in six cities (NYC, LA, Miami, Chicago, Nashville, and Seattle), bringing together over 250+ parenting influencers and local stakeholders to educate them on the tools, features, and resources we provide to support parents and teens. Attendees shared content and information from the events about our tools, features and resources, and those pieces of content reached nearly 50 million impressions. And we recently announced a series of Screen Smart workshops to help empower parents to confidently manage their teens' usage of smartphones and devices—including on Meta's platforms.

Additionally, we work closely with external groups such as ConnectSafely and Net Family News to develop resources for parents and guardians to help them have meaningful and open conversations with their teens about being online. And through our partnership with Smart Design, we conducted co-design sessions with teens and parents and consulted with experts in the US, the UK, Ireland, Brazil, Japan and India. That co-design work invites young people, parents, and experts to participate as collaborators in our design process, empowering them to provide input about how our services can meet their needs.

We also built a Family Center to help teens and families build healthy online habits. The Family Center is a central place for parents and guardians to access supervision tools and resources from leading experts. It includes an education hub where parents and guardians can access resources

from experts and review articles, videos, and tips on topics like how to talk to teens about social media. Parents and guardians can also watch video tutorials on how to use these new supervision tools. Our vision for the Family Center is to eventually allow parents and guardians to help their teens manage experiences across Meta technologies, all from one place.

Nonetheless, it can be challenging for parents to supervise the many apps that their teens may use, which is one of the reasons we support federal legislation at the app store level that would make it simpler for parents to oversee their teens' online lives. Parents want to be involved in their teen's online lives, and recent Pew research suggests that 81% of US adults support requiring parental consent for teens to create a social media account. But technology is constantly changing and keeping up with all the apps teens use can feel impossible. As an industry, we should come together with lawmakers to create simple, efficient ways for parents to oversee their teens' online experiences.

That is why we also support federal legislation that would make it simpler for parents to oversee their teens' online lives, including legislation that would require app stores to get parents' approval when teens under 16 download an app. According to a recent Morning Consult poll,[2] parents across both sides of the aisle overwhelmingly support this approach. 81% of Democratic-leaning and 79% of Republican-leaning parents back federal legislation for parental approval of teen app downloads. Over 75% of parents prefer app stores as more secure and straightforward venues for approving downloads, and a more effective method than individual app-level.

In addition to offering a simpler way for parents to approve their teens' app downloads, federal legislation needs to create standards for all apps to adhere to in areas like age-appropriate content, age verification, and parental controls.

We want to help find workable solutions and earlier this year we proposed a framework for legislation.[3] We designed this framework to create clear, consistent standards for all apps, to empower parents and guardians, and to preserve user privacy, in ways that are technologically feasible for the industry. This framework would:

- **Require app stores to get parental approval for teens under 16 to download an app.** Empowering parents to approve their teens' app downloads ensures that they oversee their teens' online experience. Placing the point of approval within the app store simplifies the process and leverages optional approval systems already offered by app stores. App stores would notify parents and request their approval when their teen wants to download an app, including Instagram.

[2] Morning Consult Survey
[3] A framework for legislation to support parents and protect teens online (January 16, 2024)

- **Require certain apps, including social media apps, to offer supervision tools for teens under 16 that parents can activate and control.** Parents should have the tools they need to guide and support their teens online. Certain apps, including social media apps, should be required to offer some form of parental supervision tools, including the ability to set daily time limits on teens' usage, see which accounts their teen is following or friends with, and more. Furthermore, apps can quickly and easily implement these tools if a parent relationship is established in the app store.

- **Require app stores to verify age and provide apps and developers with this information.** Knowing a person's age helps ensure that apps can easily place teens in the right experience for their age group, but parents and teens should not have to provide sensitive information like government IDs to hundreds of apps to verify their age. Parents already provide this information when they purchase a teen's phone and set up their teen's account. App stores have this information and not only can they ease the burden on parents by sharing it with apps, they can help ensure teens are placed in age-appropriate experiences.

- **Require industry to develop consistent age-appropriate content standards across the apps teens use.** Parents are eager to have a better understanding of the content available to their teens and to have guidelines to help them evaluate whether an app is appropriate for their child. We need broader alignment across industry on the types of content companies should consider age appropriate, as there is for other media like movies and video games. It is time we have common industry standards for what is age-appropriate that parents can rely on.

- **Establish national standards to unify the complicated patchwork of inconsistent state laws, and that apply to all apps consistently.** Parents expect consistent standards across all the apps their teens access—regardless of where their teens access or use them.

- **Require industry to develop ads targeting and delivery standards that, for example, limit the personalization of ads for teens under 16 to age and location only.** Industry standards on ad targeting and delivery can help to ensure teens see relevant ads for age-appropriate products and services in their community (e.g., a college prep course) while eliminating the ability to target this audience based on online behaviors or activity. Personalizing ads by age and location is common across industries: for example, advertisers may place relevant ads during teens' TV shows, or in magazines or newspaper sections designed for teens.

Question 7. **Concerning international law,**

a. what steps have your company and its subsidiaries taken to comply with the European Union's Digital Services Act?
b. what steps has your company and its subsidiaries taken to comply with the United Kingdom's Online Safety Act?
c. what steps has your company and its subsidiaries taken to comply with Australia's Online Safety Act?
d. if those laws create a safer, healthier online experience for kids on your platforms, do you commit to implement these changes in the United States? If not, why not?

Meta has long advocated for harmonized regulation that effectively protects people's rights online, while continuing to enable innovation. With regulation, we would welcome ambitions for greater transparency, accountability, and user empowerment.

The European Union's Digital Services Act (DSA) is an example; the DSA provides greater clarity on the roles and responsibilities of online services. The DSA came into force for Facebook and Instagram in August 2023, and for other in-scope services on February 17, 2024. As the regulation is relatively nascent, we do not yet know how it will evolve in practice.

We have been working hard since the DSA came into force to respond to these new rules and adapt the existing safety and integrity systems and processes we have in place in the areas regulated by the DSA. Our efforts include measures to increase transparency about how our systems work, and to give people more options to tailor their experiences on Facebook and Instagram. We have also established a new, independent compliance function to help us meet our regulatory obligations on an ongoing basis. As an example of our compliance with the DSA's transparency requirements, pursuant to Articles 15, 24, and 42 of the DSA, Meta published its first DSA transparency reports for Facebook and Instagram, as the designated very large online platforms of Meta Platforms Ireland Limited.

Australia's Online Safety Act is another example. We have been removing harmful content referred to us by the eSafety Commissioner since 2021 and in 2022, we responded to transparency notices as part of the Basic Online Safety Expectations notice regime, outlining how we combat child sex abuse material on our services. In 2023, the e-Safety Commissioner approved Australia's industry online safety codes as a further regulatory tool under the Online Safety Act. The codes set out the measures that online industry participants must take to enhance online protections by reducing access and exposure to certain types of harmful online material, including material promoting child sexual abuse. We remain committed to cooperating with the Office of the e-Safety Commissioner and the broader industry on the operation of these codes, any future codes, and any legal standards as they develop under the Act.

As for the United Kingdom's Online Safety Act, we are supportive of its aims. We are currently awaiting further development of the Codes of Practice and for guidance to be finalized and issued by Ofcom.

We were early supporters of creating a regulatory regime in Europe and Australia that minimizes harm effectively, protects and empowers people, and upholds their fundamental rights. Similarly, we support regulators across the globe working together to establish clear, consistent laws that adapt to ever-evolving technologies, so they can be implemented successfully by companies across our industry.

Question 8. **Last June, the Wall Street Journal reported that Instagram, "helps connect and promote a vast network of accounts openly devoted to the commission and purchase of underage-sex content." According to the report, small teams of researchers at Stanford University and the University of Massachusetts-Amherst were each able to identify these networks without any inside access to Meta's systems.**

In Meta's response to my letter following this article, Meta responded by indicating that "[b]etween 2020 and 2022, [it] dismantled 27 abusive networks, and in January 2023, [it] disabled more than 490,000 accounts for violating our child safety policies."

a. **Why were these small teams of researchers able to detect this problem while Meta, with all its resources, was not?**
b. **Does Instagram still connect and promote networks of pedophiles? What is the basis for your response?**

Preventing child exploitation is one of the most important challenges facing our industry today. That is why last year, we created a task force to address allegations about the effectiveness of our work in this area. As part of that work we reviewed existing policies; examined technology and enforcement systems we have in place; and made changes that strengthened our protections for young people, banned predators, and removed networks they use to connect with one another. Our child safety teams continue to work on additional measures. The task force focused on three areas: recommendations and discovery, restricting potential predators and removing their networks, and strengthening our enforcement. For example, we use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts, and we do not recommend teen accounts to these accounts, or vice versa. We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. We identified and removed more than 90,000 accounts from August 1, 2023 to December 31, 2023 as a result of this method.

On Instagram, potentially suspicious adult accounts are not recommended to each other in places like Explore and Reels, and are not shown comments from one another on public posts, among other things. On Facebook, we are using this technology to help better find and address certain Groups and Pages. For example, Facebook Groups with a certain percentage of members that exhibit potentially suspicious behavior will not be suggested to others in places like Groups You Should Join. Additionally, Groups whose membership overlaps with other Groups that were removed for violating our child safety policies will not be shown in Search. As we reported in December 2023, since July 1, 2023, we removed more than 190,000 Groups from Search. From July 1, 2023 to December 31, 2023, we also reviewed and removed over 21,000 Facebook Groups that violate our child safety policies.

We also hire specialists with backgrounds in law enforcement and online child safety to find predatory networks and remove them. These specialists monitor evolving behaviors exhibited by these networks—such as new coded language—to not only remove them, but to inform the technology we use to proactively find them. Between 2020 and 2023, our teams disrupted 37 abusive networks and removed more than 200,000 accounts associated with those networks.

However, online predators are determined criminals who use multiple apps and websites to target young people. They also test each platform's defenses, and they learn to quickly adapt. That is why now, as much as ever, we are working hard to stay ahead. This is also why we collaborate with industry on new programs, such as Lantern. Lantern is a program from the Tech Coalition that enables technology companies to share a variety of signals about accounts and behaviors that violate their child safety policies. Lantern participants can use this information to conduct investigations on their own platforms and take action.

c. What steps are you personally taking, if any, to make sure this never happens on Instagram or Meta's other platforms ever again?

We want everyone who uses our services to have safe, positive, and age-appropriate experiences, and we approach all our work on child safety with this in mind. We build comprehensive controls into our services, we work with parents, experts, and teens to get their input, and we engage with Congress about what else needs to be done.

We are committed to protecting young people from abuse on our services, but this is an ongoing, industry-wide challenge. As discussed in response to your Questions 8(a) and (b) above, this is a highly adversarial space that requires an industry-wide, comprehensive approach. As we improve defenses in one area, criminals shift their tactics, and we evolve our responses to address the changing threat. We will continue working with parents, experts, industry peers, and Congress to help improve child safety, not just on our services, but across the internet as a whole.

Question 9. **A December article in the Wall Street Journal reported that Meta's response is spotty—at best—when it is alerted to problem accounts and user groups. The article indicates that the Canadian Centre for Child Protection said, "a network of Instagram accounts with as many as 10 million followers each has continued to livestream videos of child sex abuse months after it was reported to the company." Other entities similarly report lengthy delays in responses to complaints about child sexual abuse material and child exploitation on Meta platforms.**

a. What is your average response time to reports of child exploitation?

As discussed in response to your Question 8, we have built sophisticated technology so we can find, remove, and report more exploitative content than any other company that reports to NCMEC. We will continue to refine our systems, and we call upon the rest of the industry to do the same.

b. In your opinion, how long is it acceptable for an Instagram account to continue livestreaming child sex abuse after it has been reported to the company?

We believe that any instance of content that violates our Child Sexual Exploitation, Abuse, and Nudity policy is one too many, which is why we continue to invest heavily in combating online child exploitation.

Over the years, we have invested heavily in sophisticated technology that helps us proactively find violating content and accounts of this kind and remove them. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in identifying certain content for human review.

Since 2019, we have also made two technologies—PDQ and TMK-PDQF—publicly available which detect identical and nearly identical photos and videos. We use PhotoDNA and other photo- and video-matching technologies that detect identical or near-identical photos and videos of known child exploitative content, and we use Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess. We also use technology to detect and remove Instagram Reels and Stories that violate our Community Guidelines, including by scanning for CSAM terms and for CSE indicators.

In August 2023 alone, we disabled more than half a million accounts on Facebook and Instagram for violating our Child Sexual Exploitation policies. And after launching a new automated enforcement effort in September, we saw five times as many automated deletions of Instagram Lives that contained nudity and sexual activity.

The overwhelming majority of people use Facebook and Instagram Live for positive purposes, like sharing a moment with friends or raising awareness for a cause they care about. When we

become aware of CSAM on Live we remove it, report it, and take action against the account responsible.

c. Why doesn't Meta suspend accounts reported for distributing CSAM while the company investigates the matter?

We work to find, remove, and report child sexual abuse material and disrupt the networks of criminals behind it. We promptly disable accounts for various violations of our child exploitation policies, such as the apparent malicious distribution of CSAM or sexual solicitation of children. As required by law, we report all apparent instances of child exploitation identified on our site from anywhere in the world to NCMEC, which coordinates with law enforcement authorities from around the world. Suspending accounts based on reports alone, as suggested by your question, would most certainly lead to overenforcement in this space and the suppression of policy compliant content and speech. Adversarial actors can and do try to exploit reporting to suppress viewpoints they disagree with, including those of policymakers, elected officials, and candidates.

d. What accountability should Meta face when it fails to respond to reports of child exploitation on its platform?

We want everyone who uses our services to have safe, positive, and age-appropriate experiences, and we approach all our work on child safety with this in mind. We build comprehensive policies and controls into our services, we work with parents, experts, and teens to get their input, and we engage with Congress about what else needs to be done. In fact, NCMEC has acknowledged Meta as an industry leader in this work and that Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs." In 2022, we made over 26 million reports between Facebook, Instagram, and WhatsApp. The rest of the industry made less than 5 million reports collectively.

We are committed to protecting young people from abuse on our services, but this is an ongoing challenge. As we improve defenses in one area, criminals shift their tactics and we evolve our responses to address the changing threat. We will continue working with parents, experts, industry peers, and Congress to try to improve child safety, not just on our services, but across the internet as a whole.

Question 10. **Meta has long submitted more CyberTips to the National Center for Missing & Exploited Children than any other platform. But in December, Meta announced it was implementing end-to-end encryption by default for messages on its Facebook and Messenger platforms. Instagram is expected to do the same soon. An assessment of the impact of this change, commissioned by Meta, said that "some of the most severe human**

rights risks of Meta's expansion of end-to-end encryption involve the use of end-to-end encrypted messaging to facilitate the sexual abuse and exploitation of children."

a. How many fewer CyberTips does Meta expect to submit to NCMEC once it implements end-to-end encryption on these platforms?

The Human Rights Impact Assessment to which your question refers concludes that encryption plays an important role in protecting human rights, supporting Meta's decision to implement end-to-end encryption across its messaging services, and offering 45 suggestions on how to address the loss of visibility into message content. It is also important to note that people overwhelmingly use Meta's services for lawful purposes. Content that contains apparent CSAM that Meta finds constitutes approximately 0.0001% of overall messages sent on Messenger.

Nevertheless, implementation of encryption on Messenger does not undercut our commitment to work with law enforcement, nor does it mean we will stop reporting harmful content to the National Center for Missing and Exploited Children (NCMEC). Indeed, we have spent more than a decade developing policies and technologies to help keep young people safe and to keep predators from attempting to use our services to connect with one another. Our comprehensive approach includes cutting-edge technology to prevent, detect, remove, and report violations of our policies that prohibit child exploitation, as well as providing resources and support to victims. We work with professionals, collaborate with industry, and support law enforcement around the world to fight the online exploitation of children. For example, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations.

In an end-to-end encrypted environment, artificial intelligence tools can help us proactively detect accounts engaged in potentially malicious patterns of behavior, a capability that helps us spot and address problems at a broad scale across our services. In addition, our machine learning technology can look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect potentially suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors. To help us respond to violations of our policies quickly, we also encourage people to report messages to us in both encrypted and unencrypted services. We have made our reporting tools easier to find and started encouraging teens to report at relevant moments, such as when they block someone.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from

scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

After the introduction of end-to-end encryption in Messenger, we expect to continue providing more reports to NCMEC than all of our peers, thanks to our industry-leading work on keeping people safe. For example, WhatsApp—which has long been encrypted—removes hundreds of thousands of accounts per month for suspected CSAM violations. In 2022, WhatsApp also made over one million reports to NCMEC, all without breaking encryption. This was significantly more than all other encrypted messaging services combined. As another data point, of the reports we made to NCMEC in the first quarter of 2021 alone, 1.2 million were generated without scanning private messages, which was more than half of the 2.3 million total reports to NCMEC from the rest of industry in all of 2021.

NCMEC has acknowledged that Meta continues to be an industry leader in this work. We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

b. Does Meta acknowledge that the change will inevitably lead to more child exploitation going undetected?

Please see the response to your Question 10(a).

c. How will Meta assess whether the adoption of encryption has made its platform more appealing to individuals committing online child sexual exploitation?

Please see the response to your Question 10(a).

Question 11. **Unlike other encrypted products that are used primarily for communication among people you already know, Facebook and Instagram are social media platforms with search and recommendation functions that connect users with strangers, including connecting adults with children. This increases the danger kids on Meta's platforms potentially face.**

This increased risk did not go unnoticed by Meta's employees. The Wall Street Journal reported in December that an engineering director at Facebook threatened to resign if Facebook moved forward with encrypting messages—which he ultimately did. Other employees made suggestions that would protect kids, such as not enabling encryption on teen accounts and not recommending minors to adults via Facebook's "People You May Know" algorithm. All of these suggestions were rejected by Facebook executives.

Why didn't Meta implement the suggestions put forward by its own employees to protect kids from the potential dangers of encrypted messaging?

Meta hires people who care deeply about these issues, and we expect them to ask questions, propose ideas, and challenge leaders of the company. It is one of the reasons we have made so much progress. While we may not adopt every proposal, we take input seriously and use it to make informed decisions. We want teens to have safe, age-appropriate experiences on our apps, including on our encrypted services. We do not believe moving to an encrypted messaging environment means sacrificing safety. To the contrary, encryption is part of how we support the privacy, safety, and security of the people who use our apps. It is already widely used by other large messaging services to help protect people's private messages and provide people with the privacy and security they expect when messaging friends and family. That is why we will continue to support encryption, while putting other features in place, as well, to help keep people safe.

We have started to roll out end-to-end-encryption on Messenger and Instagram Direct Messaging after thoughtful conversations with numerous stakeholders across the company, as well as years of consultations with leading safety and security experts across the globe. We value our employees' opinions, and we take suggestions and proposals related to safety seriously. Even when we did not take on a specific recommendation, these conversations contributed to our approach to safe encrypted experiences, which is focused on three key elements: (i) preventing potential harm in the first place; (ii) giving people ways to control their experience; and (iii)

responding to violations of our policies quickly. This approach is detailed in our whitepaper, Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging.

We have invested in and continued improving our tools and policies specifically to help young people manage interactions with adults and to reduce potential risks. For example, on Facebook and Instagram, we work to not recommend to anyone—through Facebook's "People You May Know" algorithm or otherwise—accounts we identify as exhibiting potentially suspicious behavior.[4] Specifically, we work to ensure that teens are not recommended to adult accounts exhibiting potentially suspicious behavior, and adult accounts exhibiting potentially suspicious behavior are not recommended to anyone (including to teens or other potentially suspicious adult accounts). Furthermore, in the US, accounts for people under 16 are defaulted to private, so teens can control who sees or responds to their content.

In addition, we have put in place numerous other tools to reduce potential risks related to adults connecting with teens. We restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know. We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people

[4] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

We also have developed new supervision tools that help parents manage the experiences of their teens who use our services. Parents of teens (under 16) whose accounts are enrolled in these parental controls are prompted to approve or deny their teen's requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen whose account is enrolled in these controls tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

In addition, in an end-to-end encrypted environment, we use machine learning to proactively detect accounts engaged in potentially malicious patterns of behavior. Our machine learning technology will look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect potentially suspicious activity and abuse, which is a capability that exists on Facebook and Instagram, but not on other platforms to which your question refers. For example, if an adult account repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they are not already connected to or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors.

To help us respond to violations of our policies quickly, we encourage people to report messages to us in both encrypted and unencrypted services. On Instagram, we have developed proactive safety notices that inform teens when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give them an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting potentially suspicious activity and educating people on how to take action. These notices help

people avoid scams, spot impersonations, and flag accounts that have been exhibiting potentially suspicious behavior that attempt to connect to minors.

Keeping young people safe online has been a challenge since the advent of the internet. For as long as the internet has existed, criminals have used multiple online services to target young people, testing each platform's defenses and continually adapting to a platform's countermeasures. That is why now, as much as ever, we are working hard to stay ahead of these threats by developing technology to root out predators, working with specialists dedicated to online child safety, and sharing information with our industry peers and law enforcement.

Question 12. **You described 2023 as a "year of efficiency" for Meta. In making the company more efficient, news reports indicated that Meta eliminated approximately 21,000 jobs, including many in its trust and safety teams.**

How, if at all, has Meta measured the impact of these cuts on its efforts to stop online child sexual exploitation? Please provide any studies, reports, summaries, or data that relate to the impact of these cuts on Meta's efforts to stop online child exploitation.

It is incorrect to characterize the restructuring efforts referred to in your question as having an outsized impact on trust and safety teams. After years of growth, Meta implemented a company-wide restructuring plan focused on flattening our organization. The goal of these efforts was to make the company faster, leaner, and more efficient. To be clear, these restructuring efforts did not change the commitment we have to our ongoing integrity efforts, including our ongoing child safety efforts. We have brought teams together to think across a number of key issues. For example, our Global Operations team now works more closely with our integrity team, and we have consolidated certain support teams from different areas across the company.

To be clear, we absolutely remain committed to our work keeping people safe on our services. Even with the targeted changes, we continue to have about 40,000 people focused on overall safety and security efforts. Finding efficiencies in our work has been a focus for years. We will continue to hire across security and integrity teams to support our industry-leading work in the most efficient and effective manner possible.

Question 13. **In 2021, Instagram's Well-Being Team emailed senior executives raising concerns that the company's investment in staff was insufficient to address problematic use of the platform by teens. A few days later, Nick Clegg, Meta's President of Global Affairs, emailed you to say that this lack of investment was delaying innovations to protect children. A few months later, Meta's Responsible Innovation Team was disbanded, positions were cut**

from Instagram's Well-Being Team, and more than 100 positions related to trust, integrity, and responsibility were eliminated.

a. Why did you reject the recommendations of your own staff to bolster resources addressing teen well-being?

At Meta, we hire people who care deeply about well-being, and we expect them to ask questions, propose ideas, and challenge the leaders of the company. This open dialogue is how we have made so much progress. While we may not adopt every proposal, we take input seriously and make informed decisions about the best way forward.

We are committed to helping teens have safe and positive experiences on our services, and we have devoted enormous resources to put in place tools, features, and policies to support teens and their parents. For example, we have developed more than 50 tools and features designed to support teens and their families—ideas that were brought forth by our internal experts and external advisors.

It is also not accurate that the issues raised in the email were rejected or ignored. For example, we have a number of new features, including parental supervision tools that let teens work with their parents to set daily limits for the total time that teens can spend on our apps, Take A Break notifications, which show full-screen reminders to leave the Instagram app, Quiet Mode, which turns off notifications at night, and Nudges, which include alerts that notify teens that it might be time to look at something different if they have been scrolling on the same topic for a while.

We have every incentive to provide safe platforms that people enjoy and that advertisers want to use. This is core to our company values. That is why we have invested more than $20 billion in safety and security since 2016—including approximately approximately $5 billion in 2021 and $6 billion in 2022—and we will never stop working on these issues.

For more on our continued commitment to investing in efforts to protect our youngest users on our platforms, please see the response to your Question 12.

b. What accountability should Meta face if it fails to deploy sufficient resources to protect children from online child sexual exploitation?

We work hard to provide support and controls to reduce potential online harms, and it is important to us at Meta that our services are positive for everyone who uses them. Meta has around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone. We go beyond legal requirements and use sophisticated technology to proactively seek out abusive material, and as a result, we find and report more inappropriate content than anyone else in the industry. As the National Center for Missing and Exploited Children (NCMEC) recently put it, Meta goes "above

and beyond to make sure that there are no portions of their network where this type of activity occurs." Still, no matter how much we invest or how effective our tools are, there is always more to learn and more improvements to make. When it comes to industry accountability, we remain ready to work with members of this Committee, the industry, and parents to strengthen our services and make the internet safer for everyone.

Question 14. **Meta recently announced its support for kids' online safety legislation centered around "requir[ing] app stores to get parents' approval whenever their teens under 16 download apps."**

Less than two years ago, Meta opposed the idea of putting more power in the hands of app stores. The company submitted comments to the National Telecommunications and Information Administration in May 2022 complaining about the dominance of Apple's App Store. The company claimed it was "at the mercy of Apple policies that gate our access to people, creators, and businesses who enjoy and value our mobile applications."

Now, Meta is trying to codify Apple's gatekeeper status.

What explains Meta's change in its position?

Meta's comments before the National Telecommunications and Information Administration (NTIA) were in the context of an NTIA inquiry into the state of competition of the mobile app ecosystem. Meta comments provided an overview of competition in the mobile app ecosystem and focused on actions at the operating system and app store level that limited growth, competition, and innovation by third-party developers in that specific context.

Our position in our NTIA comments is consistent with our position today. We support federal legislation that requires age verification and parental approval at the app store. We have ongoing concerns about Apple's exercise of power in the mobile app ecosystem, which are heightened in the context of youth well-being in light of reports of Apple's smartphone market share among US teenagers approaching 90%. To ensure youth well-being legislation does not enable Apple to further entrench its dominant position and find new ways to exclude competition, Meta also supports competition safeguards that Congress should enact as part of any youth well-being legislation that would prohibit app store operators from engaging in certain forms of discriminatory self-preferencing.

Our position on parental approval for app downloads is focused on empowering parents. We recognize that parents want to be involved in their teen's online lives, and recent Pew research suggests that 81% of US adults support requiring parental consent for teens to create a social

media account. But technology is constantly changing and keeping up with all the apps teens use can feel impossible.

The best way to help support parents and young people is a simple, industry-wide solution where all apps are held to the same, consistent standard. Parents should approve their teen's app downloads, and we support federal legislation that requires app stores to get parents' approval whenever their teens under 16 download apps. With this solution, when a teen wants to download an app, app stores would be required to notify their parents, much like when parents are notified if their teen attempts to make a purchase. Parents can decide if they want to approve the download. They can also verify the age of their teen when setting up their phone, negating the need for everyone to verify their age multiple times across multiple apps.

This way parents can oversee and approve their teen's online activity in one place. They can ensure their teens are not accessing adult content or apps, or apps they just do not want their teens to use. And where apps like ours offer age-appropriate features and settings, parents can help ensure their teens use them.

Question 15. **Last month, a group at Stanford found that a popular dataset used to train generative AI contained hundreds of images of CSAM.**

You have said that Meta is training its next generative AI model, Llama 3, right now, and that your ambition is for Meta's AI to eventually be "the leading models in the industry."

As Meta makes strides to become a leader in AI, what is Meta doing to ensure its AI isn't trained using datasets that include illegal CSAM content?

We work to minimize the possibility of illegal child sexual abuse material being used to train our AI models. We also work with experts and industry partners to help prevent Generative AI models from being used to harm children, and we are routinely testing our models to help our AI features provide experiences that are safer and more helpful for young people.

We also strive to use a number of protections in our generative AI, including:

- **Training Our Model to Recognize Exploitative Queries**: We are training our models to recognize different types of queries, including those related to child exploitation or sexualization, and to not provide a response to certain queries which may be harmful or illegal, including child exploitative materials.
- **Continual Testing**: Dedicated teams work with internal child safety experts and use our institutional knowledge of child safety risks online to test our models with terms and

prompts that may be used by those seeking to harm children, allowing us to identify and address inappropriate responses.

- **Removing Violating Content from Responses**: Building on our long-standing investment in technology that helps to proactively find and remove child exploitative content, we have implemented new technology into our models that works to prevent such content from responses before they are shared with people, in the event the model were to initially generate a response. For example, if someone prompts our AI to create content that could exploit or harm children, our proactive technology works to scan responses and prevent those that may relate to child exploitative content from being shown.
- **Providing Feedback on Responses**: We have developed feedback tools so people can flag responses that they perceive to be unsafe or offensive, and we will use this feedback to continue training the models and improve our ability to restrict our AI from providing such responses.

Question 16. **Snapchat's disappearing message features has made it a platform of choice for those looking to engage in sextortion. Yet, in December, Meta announced it was introducing its own version of a disappearing message feature on Facebook and Messenger. In its announcement, the company acknowledged that disappearing messages provide a false sense of security, as the recipient can save an image by simply taking a screen shot. While, someone using vanish mode on Instagram or Messenger will be told if the recipient takes a screenshot of the message, for a teen who sends a sexually-explicit photo, that's already too late.**

What additional steps is Meta taking to prevent disappearing messages from putting more kids at risk of sextortion?

Having a personal intimate image shared with others can be devastating, especially for young people. It can feel even worse when someone threatens to share that image if a person does not give more photos, sexual contact, or money—a crime in most jurisdictions, commonly referred to as sextortion.

At Meta, we take a multi-faceted approach to combat sextortion. These efforts include (i) strict policies against content or activity that sexually exploits or endangers children, including sextortion; (ii) human and machine detection and enforcement, including automated rules that detect and action at scale accounts committing financial sextortion; (iii) education and safeguards to help prevent suspicious adult accounts from finding or interacting with teens on our apps; and (iv) proactive investigatory work, including targeted investigations and removal of violating accounts to disrupt networks of bad actors attempting to exploit or financially extort minors,

and—when appropriate—reporting them to the National Center for Missing and Exploited Children (NCMEC). These efforts are described in more detail below, as well as Meta's position on disappearing messaging.

We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. We work to prevent this content, as well as inappropriate interactions between young people and suspicious accounts attempting to take advantage of them. We also prohibit behavior that exploits people, including sharing or threatening to share someone's intimate images. In addition, impersonation is one way criminals gain the trust of their sextortion victims, which is one reason why our policies prohibit it. This helps address the problem at the root and prevent downstream harms, like sextortion. We have invested heavily in strengthening our technology to keep fake accounts off Facebook and Instagram and we cooperate with law enforcement and respond to lawful information requests in prosecutions of scammers.

We encourage people across our services to report both encrypted and unencrypted messages if they have a concern about them, and we have a dedicated reporting option to use if someone is sharing private images. While disappearing messages on Messenger are only available for end-to-end encrypted conversations, people can still report them if they receive something inappropriate. Additionally, if we detect that someone screenshots a disappearing message, we notify the sender. People can report nude or sexual photos or videos of themselves or threats to share these images or videos to our apps or technologies to prevent them from being reshared. Our teams review reports 24/7 in more than 70 languages. We have articles in our Help Center that help people understand how to report this activity on Facebook,[5] Instagram,[6] and Messenger.[7]

Moreover, we have specialized teams working on combating sextortion. These teams are constantly working to understand the unique combinations of on-platform behaviors used by criminals seeking to exploit our services. These criminals often impersonate others, including minors, to gain the trust of their victims and in violation of our policies. We build automation rules that allow us to detect and action—at scale and with high-precision—accounts committing financial sextortion. Our teams continue to work on new solutions to address sextortion industry-wide, including by developing new ways to identify people potentially engaging in sextortion and thwarting their efforts.

Our dedicated teams investigate and remove these criminals and report them to authorities, including law enforcement and NCMEC, when appropriate. We work with partners, like

[5] How do I report an abusive photo on Facebook? | Facebook Help Center
[6] How to Report Things | Instagram Help Center
[7] Reporting Conversations | Messenger Help Center (facebook.com)

NCMEC and the International Justice Mission, to help train law enforcement around the world to identify, investigate and respond to these types of cases. We have developed a streamlined online process through which we accept and review all legal requests from law enforcement. If we have reason to believe that a child is in immediate or imminent danger, we may proactively refer a case to local law enforcement (as well as report it to NCMEC) to help safeguard the child.
We work to protect people from sextortion by helping to prevent unwanted contact across our apps and in our messaging services, especially between adults and teens. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists, which can be used as a lever by people trying to sextort others. We do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We also restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts.[8] Specifically, we work to ensure that teens are not recommended to potentially suspicious adult accounts, and potentially suspicious adult accounts are not recommended to anyone (including to teens or other potentially suspicious adult accounts). We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help

[8] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

people avoid scams, spot impersonations, and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Because this is an industry-wide concern, we also direct people to various tools to use if people have nude or sexual photos or videos to help prevent them from being shared or reshared online. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Finally, with respect to disappearing messages generally, we believe that people should have simple, intimate places where they have clear control over who can communicate with them and confidence that no one else can access what they share. In addition, people should be comfortable being themselves, and should not have to worry about what they share coming back to hurt them later. So we will not keep messages or stories around for longer than necessary to deliver the service or longer than people want them.

For more information about our work combating sextortion and intimate image abuse, please see our dedicated page in Safety Center, linked here: https://about.meta.com/actions/safety/topics/bullying-harassment/ncii.

Question 17. **In your written testimony, you state that "in the last 8 years [Meta has] introduced more than 30 different tools, resources, and features to help parents and teens."**

Please provide an itemized list of those tools, specifying when they were launched and describing how they were assessed for efficacy before and after they were implemented.

We have built numerous tools, features and resources that help teens have safe, positive experiences. As reflected in the preceding link, we have reproduced a timeline below with some of the tools, features and resources we have developed across our apps and technologies. We provide more information about these tools and features, and how they work, in our Instagram Parent Guide and our Family Center, and additional resources on how to create supportive online experiences, such as our Education Hub for Parents and Guardians, our Safety Center, and our Bullying Prevention Hub.

- **October 2010:** Instagram launches with the blocking feature.
- **September 2016:** We gave people the option to swipe to delete comments that they found inappropriate on Instagram.

- **September 2016:** We launched our comment keyword filter on Instagram, allowing people to filter out offensive or inappropriate comments
- **December 2016:** We gave people the option to turn off comments on Instagram.
- **December 2016:** We launched anonymous reporting of accounts that may be struggling with their mental health, and directing those accounts to resources on Instagram.
- **March 2017:** We added the ability for people to connect with crisis support partners like Crisis Text Line and the National Eating Disorders Association on Messenger.
- **June 2017:** We launched our offensive comment filter control, allowing people to automatically hide certain offensive comments on Instagram. We later expanded the offensive comment filter to include terms related to bullying and harassment.
- **September 2017:** We gave people the option to choose who can comment on their posts on Instagram.
- **September 2017:** We gave people the option to file an anonymous report of potential self-injury in Live, and provided resources to those affected on Instagram.
- **August 2018:** We launched an activity dashboard, which included a daily reminder and a new way to limit notifications on Instagram and Facebook.
- **July 2019:** We began showing Comment Warnings on Instagram to prompt people to reconsider posting comments that may be hurtful. We later expanded this feature to include an additional, stricter warning when people repeatedly try to post potentially offensive comments, and more details about what could happen if they choose to proceed.
- **October 2019:** We launched Restrict, a feature that allows people to control their Instagram experience, without notifying people who may be attempting to target them.
- **December 2019:** We began showing Caption Warnings on Instagram, to prompt people to reconsider posting images and captions that may be offensive or hurtful.
- **May 2020:** We launched the ability for people to delete multiple comments at once.
- **May 2020:** We launched the ability for people to block or Restrict multiple accounts at once. We later launched "multi-block," an option for people to both block specific accounts and preemptively block new accounts that someone may create to target them.
- **May 2020:** We gave people the option to pin comments, to give people an easy way to amplify and encourage positive interactions.
- **May 2020:** We gave people the option to manage who can tag and mention them on Instagram, to help protect themselves from bullies who may try to target them in this way.

- **November 2020:** We added a message at the top of all search results when people searched for terms related to suicide or self-injury on Instagram, pointing to resources.
- **February 2021:** We launched expert-backed resources when someone searches for eating disorders or body image-related content, and in May we launched a dedicated reporting option for eating disorder content.
- **March 2021:** We restricted people over 19 years old from sending private messages to teens who do not follow them.
- **March 2021:** We began using safety notices to encourage teens to be cautious in conversations with adults they are already connected to.
- **April 2021:** We launched our Hidden Words tool to give people the option to filter direct message requests containing certain offensive words, phrases, and emojis.
- **May 2021:** We gave people the ability to hide public like counts, to give them more control over their experience.
- **July 2021:** We began limiting potentially suspicious adult accounts from finding and following teens in places like Reels, Explore, or "Suggested for You."
- **July 2021:** We announced default private account settings for teens under 16 when they sign up for Instagram, as well as notifications encouraging existing teens under 16 to switch to a private account.
- **July 2021:** We launched our Sensitive Content Control, which allowed people to decide how much sensitive content shows up in Explore. We later began defaulting all teens under 16 into the "Less" setting in Sensitive Content Control on Instagram to make it more difficult for them to come across potentially sensitive content in Search, Explore, and Hashtag Pages, Reels, Feed Recommendations, and Suggested Accounts.
- **August 2021:** We launched our "Limits" tool, which allows people to automatically hide comments and direct message requests from people who do not follow them, or who only recently followed them.
- **December 2021:** We launched "Take A Break" to empower people to make informed decisions about how they are spending their time. To make sure that teens were aware of this feature, we showed them notifications suggesting they turn these reminders on.
- **December 2021:** We began restricting people from tagging or mentioning teens that do not follow them, or from including their content in Reels Remixes or Guides when they first join Instagram.

- **February 2022:** We launched "Your activity," which allows people to bulk manage their content and interactions, review their history, and download their information.
- **February 2022:** We introduced Personal Boundary for Horizon Worlds and Horizon Venues, preventing avatars from coming within a set distance of each other and making it easier to avoid unwanted interactions. Personal Boundary is automatically turned on for everyone in Horizon Worlds.
- **March 2022:** We introduced VR Parental Supervision Tools on Quest and we launched Family Center and Parental Supervision Tools on Instagram. Initially Instagram's supervision tools allowed parents to work with teens to:
    - View how much time their teens spend on Instagram and set time limits.
    - Set specific times during the day or week to limit their teen's use of Instagram.
    - Be notified when their teen chooses to report an account or post, including who was reported and the type of content.
    - View what accounts their teens follow and the accounts that follow them.
- **March 2022:** We announced Favorites and Following, two new ways for people to choose what they see in their Instagram Feed, including giving people the option to see their feeds in chronological order.
- **May 2022:** We launched the ability for parents to lock teens out of their apps on the Quest platform.
- **June 2022:** We brought more Parental Supervision Tools to Quest headsets, allowing parents to work with teens to:
    - Approve their teen's download or purchase of an app.
    - Block specific apps that may be inappropriate for their teen.
    - Receive "Purchase Notifications," alerting them when their teen makes a purchase in VR.
    - View headset screen time from the Oculus mobile app, so they will know how much time their teen is spending in VR.
    - See a list of their teen's friends.
    - Limit a teen's ability to use their Quest with a PC or sideload apps not available on the Quest store.

- **June 2022:** We introduced Voice Mode in Horizon Worlds, which allows people to choose how they hear people whom they do not know. When Voice Mode is turned to the garbled voices setting, the voices of non-friends sound like unintelligible, friendly sounds.
- **June 2022:** We updated Parental Supervision Tools on Instagram to include more options for parents to work with teens to,:
    - Set specific times during the day or week when they would like to limit their teen's use of Instagram.
    - See more information when their teen reports an account or post, including who was reported, and the type of report.
- **June 2022:** We launched new Nudges for teens on Instagram that encourage them to switch to a different topic if they are repeatedly looking at the same type of content on Explore.
- **June 2022:** We introduced new ways to verify peoples' age on Instagram, including privacy-preserving selfie videos.
- **July 2022:** We introduced new tools that allow parents to work with their teens to enable and disable social features for teens that they are supervising in Quest, including disabling the ability for their teens to send or receive chat messages.
- **October 2022:** We began nudging people to be kind in direct message requests, to discourage offensive or inappropriate direct messages.
- **November 2022:** We began prompting teens to report accounts to us after they block someone.
- **November 2022:** We began defaulting teens under the age of 16 (or under 18 in certain countries) into more private settings when they join Facebook, and encouraged teens already on the app to choose these more private settings.
- **January 2023:** We began giving teens more ways to manage the types of ads they see on Facebook and Instagram with Ad Topic Controls.
- **January 2023:** We launched Quiet Mode, a feature to help people focus and to encourage them to set boundaries with their friends and followers. We prompt teens to turn on Quiet Mode when they spend a specific amount of time on Instagram late at night.
- **January 2023:** We made updates to give people more control over the content they see on Instagram. First, we gave people the option to choose to hide multiple pieces of content in Explore at one time. When people select "Not interested" on a post seen in

Explore, we aim to avoid showing them this kind of content in other places where we make recommendations, such as Reels and Search. We also began allowing people to customize their recommendations with keywords. People can add a word or list of words, emojis or hashtags that they want to avoid—like "fitness" or "recipes"—and we will work to no longer recommend content with those words in the caption or the hashtag.

- **February 2023:** Meta and NCMEC launched Take It Down, a tool to help prevent the spread of young people's intimate images.
- **April 2023:** We brought Parental Supervision Tools to Horizon Worlds, allowing parents to work with their teens to:
    - See, adjust, and lock safety features like voice mode and personal boundary.
    - See who their teen follows and who follows their teen.
    - See which apps their teen has used and how much time they have spent in Meta Quest and Worlds in the past seven days.
    - Give permission to allow or block their teen from using apps, including Worlds.
- **April 2023:** We introduced a new tool, the Meta Quest Browser Website Category Filter, to help parents and guardians work with their teens to manage what their teen can access and view in the Meta Quest Browser.
- **June 2023:** We brought Parental Supervision Tools to Messenger, allowing parents to work with their teens to:
    - View how much time their teen spends on Messenger.
    - View and receive updates on their teen's Messenger contacts list, as well as their teen's privacy and safety settings.
    - Get notified if their teen reports someone (if the teen chooses to share that information).
    - View who can message their teen (only their friends, friends of friends, or no one) and see if their teen changes this setting.
    - View who can see their teen's Messenger stories and get notified if these settings change.
    - We later added additional features, including: giving parents the ability to set scheduled breaks and to view their teens' blocked contacts.

- **June 2023:** We began requiring people to send an invite to get their permission to connect in direct messages. We limit these message request invites to text only, so people cannot send any photos, videos, or voice messages, or make calls, until the recipient has accepted the invite to chat. These changes mean people will not receive unwanted photos, videos, or other types of media from people they do not follow.
- **June 2023:** We began showing teens a notification when they spend 20 minutes on Facebook, prompting them to take time away from the app and set daily time limits.
- **October 2023:** We gave people the option to manually hide comments, to give them even greater control over comments that they may find upsetting or unwelcome, in addition to our Hidden Words tool.
- **November 2023:** We brought parental supervision tools to Facebook, allowing parents to oversee things like:
    - The amount of time their teen spends on Facebook.
    - To schedule breaks for their teens and access expert resources on managing their teens' time online.
- **January 2024:** We began hiding more types of age-inappropriate content for teens on Instagram and Facebook.
- **January 2024:** We began hiding more results in Instagram Search related to suicide, self-harm, and eating disorders. Now, when people search for terms related to suicide, self-harm, and eating disorders, we will start hiding these related results and will direct them to expert resources for help.
- **January 2024:** We began prompting teens to update their privacy settings on Instagram in a single tap with new notifications.
- **January 2024:** We launched new nighttime nudges that show up when teens have spent more than ten minutes on a particular Instagram surface (i.e., Reels or Instagram Direct Message) late at night. They will remind teens that it is late, and encourage them to close the app.
- **January 2024:** We announced stricter default message settings for teens under 16 (under 18 in certain countries), meaning only people they follow or people they are already connected to can message them or add them to group chats.
- **January 2024:** Building on Instagram's existing parental supervision tools, parents using supervision will now be prompted to approve or deny their teen's (under 16) requests to

change their default safety and privacy settings to a less strict state—rather than just being notified of the change.

- **February 2024:** Meta has worked with the National Center for Missing and Exploited Children (NCMEC) to expand Take It Down to more countries and languages, allowing more teens to take back control of their intimate imagery and help protect themselves from scammers. We also partnered with Thorn to develop updated tips for teens—and parents and teachers—on what to do if they are affected by scammers who seek to exploit their intimate imagery.

A large majority of teens keep their default settings. For example, among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians and teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

Questions from Senator Graham

Question 1. **Do you support S. 1207, the bipartisan EARN IT Act? Why or why not?**

As a general matter, we would support a bill requiring common industry standards for protecting children online. However, we believe that it is important to make sure any bill intending to establish such standards protects encryption. As our lives move more and more online, we believe it is critical to preserve a space for private conversations where people can have the freedom to be themselves and share their most personal thoughts with loved ones.

We also think it is important that the bill not undermine the purposes of Section 230. At a high level, Section 230 does two things. First, it encourages free expression by barring claims against online services for publishing third-party speech. Without Section 230, online services could potentially be held liable for everything people say. Without this protection, such services may be likely to remove more content to avoid legal risk and may be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows online services to remove certain objectionable content. Without Section 230, such services could face liability, for example, for removing bullying and harassment content.

Meta has long been supportive of updating Section 230, for example, to ensure that it separates good actors from bad, by making sure that companies cannot hide behind Section 230 to avoid responsibility for intentionally facilitating illegal activity on their services. We understand that people want to know that companies are taking responsibility for combating harmful content—especially illegal activity—on their online services. They want to know that when such services remove content, they are doing so fairly and transparently. Updating Section 230 is a significant decision. It is important that any changes to the law do not prevent new companies or businesses from being built, because innovation in the internet sector brings real benefits to all Americans, as well as to billions of people around the world.

We look forward to continuing engagement with your office on this bill.

Question 2. **What measures are you taking to prevent and address sextortion, including financial sextortion, on your companies' platforms?**

a. What methods are in place to detect and disrupt this type of abuse in real time?

Having a personal intimate image shared with others can be devastating, especially for young people. It can feel even worse when someone threatens to share that image if a person does not give more photos, sexual contact, or money—a crime in most jurisdictions, commonly referred to as sextortion.

At Meta, we take a multi-faceted approach to combat sextortion. These efforts include (i) strict policies against content or activity that sexually exploits or endangers children, including sextortion; (ii) human and machine detection and enforcement, including specialized teams focused on combating sextortion and automated rules that detect and action at scale accounts; (iii) proactive investigatory work, including targeted investigations and removal of violating accounts to disrupt networks of bad actors attempting to exploit or financially extort minors, and—when appropriate—reporting them to the National Center for Missing and Exploited Children (NCMEC); (iv) safeguards to help prevent suspicious adult accounts from finding or interacting with teens on our apps, including parental controls; and (v) provide education and awareness resources to those who may had their intimate images shared online. These efforts are described in more detail below.

We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. We work to prevent this content, as well as inappropriate interactions between young people and suspicious accounts attempting to take advantage of them. We also prohibit behavior that exploits people, including sharing or threatening to share someone's intimate images. In addition, impersonation is one way criminals gain the trust of their sextortion victims, which is one reason why our policies prohibit it. This helps address the problem at the root and prevent downstream harms, like sextortion. We have invested heavily in strengthening our technology to keep fake accounts off Facebook and Instagram and we cooperate with law enforcement and respond to lawful information requests in prosecutions of scammers.

We have specialized teams working on combating sextortion. These teams are constantly working to understand the unique combinations of on-platform behaviors used by criminals seeking to exploit our services. We build automation rules that allow us to detect and action—at scale and with high-precision—accounts committing financial sextortion. Our teams continue to work on new solutions to address sextortion industry-wide, including by developing new ways to identify people potentially engaging in sextortion and thwarting their efforts.

In addition, our dedicated teams investigate and remove these criminals and report them to authorities, including law enforcement and NCMEC, when appropriate. We work with partners, like NCMEC and the International Justice Mission, to help train law enforcement around the world to identify, investigate and respond to these types of cases. We have developed a streamlined online process through which we accept and review all legal requests from law enforcement. If we have reason to believe that a child is in immediate or imminent danger, we may proactively refer a case to local law enforcement (as well as report it to NCMEC) to help safeguard the child.

We also work to protect people from sextortion by preventing unwanted contact across our apps and in our messaging services, especially between adults and teens. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists, which can be used as a lever by people trying to sextort others. We also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We also restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We have developed ways to help people control their own experience. For example, people can choose who can message them, and can block anyone they do not want to hear from. People can report nude or sexual photos or videos of themselves or threats to share these images or videos to our apps or technologies to prevent them from being reshared. Our teams review reports 24/7 in more than 70 languages. We have articles in our Help Center that help people understand how to report this activity on Facebook,[9] Instagram,[10] and Messenger.[11]

Finally, we use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts.[12] Specifically, we work to ensure that teens are not recommended to potentially suspicious adult accounts, and potentially suspicious adult accounts are not recommended to anyone (including to teens or other potentially suspicious adult accounts). We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially

[9] How do I report an abusive photo on Facebook? | Facebook Help Center
[10] How to Report Things | Instagram Help Center
[11] Reporting Conversations | Messenger Help Center (facebook.com)
[12] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations, and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Because this is an industry-wide concern, we also direct people to various tools to use if people have nude or sexual photos or videos to help prevent them from being shared or reshared online. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Anyone seeking support and information related to sextortion can visit our education and awareness resources, including the Stop Sextortion resources, developed with Thorn. These resources include immediate actions parents and teens can take if users are experiencing sextortion, as well as expert tips for teens, parents and guardians, and information for parents on how to talk to their teens about intimate images. We also worked with Thorn and their NoFiltr brand to create and promote educational materials that reduce the shame and stigma surrounding intimate images, and empower teens to seek help and take back control if they have shared them or are experiencing sextortion.

We also need Congress to pass legislation requiring operating-system level age verification requirements. That would allow services like Instagram to more quickly identify suspicious behavior, such as adults pretending to be minors, and remove them from the app entirely before they can even make contact with a teen—in addition to the work we have already been doing to prevent this contact. This also allows parents to oversee and approve their teen's online activity in one place. When a teen wants to download an app, app stores would be required to notify their parents. Where apps like ours offer age-appropriate features and settings, parents can help their teens use them. Until then, we require people to provide their age when signing up for accounts on our services, which helps us to provide teens with age-appropriate experiences.

For more information about our work combating sextortion and intimate image abuse, please see our dedicated page in Safety Center, linked here: https://about.meta.com/actions/safety/topics/bullying-harassment/ncii.

***Question 3*. Please provide the committee statistics on how long it takes your company to respond to various types of legal process from law enforcement?**

We work with law enforcement, and deeply respect and support the work agencies do to keep us safe. The amount of time it takes to respond to certain legal process depends on a variety of factors. In all cases, we carefully review, validate, and respond to law enforcement requests, and we prioritize emergency situations, including terrorism and child abuse. We also reach out to law enforcement when we see a credible threat of imminent offline harm, contacting federal, state, or local law enforcement depending on the specific circumstances of a threat.

We have specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations review potentially violating content and report findings to the National Center for Missing and Exploited Children (NCMEC). The reports to NCMEC include content from around the world, and in turn, NCMEC works with US federal, state, and local law enforcement, as well as law enforcement globally, to find and help victims.

With respect to our cooperation with law enforcement, we have developed a streamlined online process through which we accept and review all legal requests from law enforcement. We expedite requests pertaining to child safety, along with other emergency situations. We have a team dedicated to engaging with NCMEC, Child Exploitation and Online Protection Command, Interpol, the FBI, Homeland Security Investigations and numerous other local, federal, and international law enforcement organizations and departments to help them have the information and training needed to make the best use of this process and support efforts to improve process. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC to help safeguard the child.

We dedicate significant resources to addressing the concerns of law enforcement authorities and ensuring the timely processing of legal requests. We have law enforcement response teams available around the clock to respond to emergency requests. From January to June 2023 alone, we produced at least some data pursuant to legal process in more than 87% of requests, and produced data in over 77% of emergency disclosure requests.

Question 4. **Do you notify your users when law enforcement serves subpoenas/summons for subscriber information and specifically requests not to notify the subscriber/user?**

a. **If you notify the subscriber, how long do you wait until notification goes out?**
b. **Are you aware that by notifying the subscriber about a law enforcement subpoena for their subscriber information that you are jeopardizing critical evidence that could be erased before law enforcement can serve warrants?**
c. **Would your company agree to a 90-day non-disclosure to subscribers to allow law enforcement ample time to secure proper legal process?**

Our policy is to notify people who use Facebook and Instagram of requests for their information prior to disclosure, unless we are prohibited by law from doing so or in exceptional circumstances, such as child exploitation cases, emergencies or when notice would be counterproductive. Law enforcement officials who believe that notification would jeopardize an investigation can, and often do, obtain an appropriate court order or other appropriate process establishing that notice is prohibited; we also comply with such process. Furthermore, our systems also allow for law enforcement agencies to continually submit additional court orders to extend non-disclosure orders that will be expiring in the near future.

Question 5. **Do you actively seek out and incorporate feedback and insight from survivors of online sexual exploitation to improve your trust and safety policies and practices and to prevent and disrupt child sexual abuse material (CSAM) production and distribution on your platform? Can you provide examples?**

a. If not, please explain.

We consult with a number of external experts and partners—including survivors and survivor organizations—as we work to provide people with a safe and positive experience on our services. As further described below, this includes other members of the technology industry, nonprofits, law enforcement, civil society organizations, and academics with relevant experience.

More specifically, we incorporate feedback from survivors in a number of ways, including collaborating with organizations who work with survivors in a safe, trauma-informed way, and meeting with survivors at conferences hosted by various stakeholders. Meta also provides funding to NCMEC's free service to help survivors and families impacted by online sexual exploitation, and maintains a dedicated reporting channel where NCMEC staff are able to flag non-CSAM posts and profiles that threaten or otherwise identify CSAM survivors for our review and action.

In addition to these engagements with survivors, to combat child exploitation both on and off our platforms, we work with child safety professionals to help us understand evolutions in coded language and to identify new and evolving terms, phrases, slang, and emojis that could be used in an attempt to evade our detection systems and bypass our policies. Our teams use these signals and technology to proactively find new trends, misspellings and spelling variations of this language, as well as terms and phrases related to child exploitation, that we can input into our systems to proactively find and disrupt efforts to evade our protections. We also work with these professionals and organizations to build various interventions and resources, including but not limited to our search interventions, safety notices, reporting flows and safety education campaigns. We have also worked with child safety researchers to conduct collaborative research to improve child safety protections on our platforms.

Our collaborative work to address child safety does not stop with improving our own services. We also are deeply committed to improving the entire ecosystem and have engaged with child safety nonprofits and academic researchers to complete child safety research with fieldwide impact. Our efforts with these professionals also include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, among other things. Additionally, to help safety stakeholders identify and respond to the most high priority reports at NCMEC, we funded and helped rebuild their case management tool to ensure investigators can get to the most important cases quickly.

Question 6. **During our hearing, you testified that you collaborate with parents and parent organizations to create mechanisms to keep children safe online. Please elaborate and cite examples of your company's work with non-employee parents and parent organizations.**

We want parents to have information to help their teens have a safe and positive experience on our services, and constantly work to make sure parents know about our parental control features. Below we discuss our collaboration with parents and external research organizations.

We reach parents in a variety of ways, including through our Family Center, advertising campaigns, in-app promotion, our work with safety partners, and events with parents. We work closely with groups like ConnectSafely, ParentZone and Net Family News to develop resources for parents and guardians to help them have meaningful and open conversations with their teens about being online. For example, our Family Center includes an education hub where parents and guardians can access resources from experts and review helpful articles, videos, and tips on topics like how to talk to teens about social media. Parents can also watch video tutorials on how to use the new supervision tools available on Instagram today. And the Meta Quest Parent Education Hub includes a guide to our VR parental supervision tools from ConnectSafely to help parents discuss virtual reality with their teens. In the US, we have collaborated with The Child Mind Institute and ConnectSafely to publish a Parents Guide. It includes the latest safety tools and privacy settings, as well as a list of tips and conversation starters to help parents navigate discussions with their teens about their online presence. And in our Safety Center, we provide co-branded resources for parents from our collaboration with expert organizations.

As part of our work developing Messenger Kids, in addition to our research with thousands of parents, we engaged with over a dozen expert advisors in the areas of child development, online safety and children's media and technology who helped inform our approach. We have also had conversations around topics such as responsible online communication and parental controls with organizations like National PTA and Blue Star Families, where we heard firsthand how parents and caregivers approach raising children in today's digitally connected world.

We have also launched TTC Labs, a global co-design program, that invites parents, young people, and experts to participate as collaborators in our design process, empowering them to have their say and ensuring our products meet their needs. And through our partnership with Smart Design, we conducted co-design sessions with parents and teens, and consulted with experts in the US, the UK, Ireland, Brazil, Japan and India, empowering them to provide input about how our services can meet their needs.

Additionally, in 2023, we hosted "Screen Smart" events in six cities (NYC, LA, Miami, Chicago, Nashville, and Seattle), bringing together over 250+ parenting influencers and local stakeholders to educate them on the tools, features, and resources we provide to support parents and teens. Attendees shared content and information from the events about our tools, features and resources, and those pieces of content reached nearly 50 million impressions. And we recently announced a series of Screen Smart workshops to help empower parents to confidently manage their teens' usage of smartphones and devices—including on Meta's platforms. Our ads encouraging parents to use our youth well-being tools and features were seen more than one billion times by tens of million people in the United States since August 2022.

With respect to the research community, we have a global team responsible for helping to ensure that Meta remains a leader in online safety. We employ and work with researchers from backgrounds that include clinical psychology, child and developmental psychology, pediatrics research, public health, bioethics, education, anthropology, and communication. We also collaborate with top scholars to navigate various complex issues, including those related to well-being for people on Facebook and Instagram.

Additionally, Meta awards grants to external researchers to help us better understand how experiences on Facebook and Instagram relate to the safety and health of our community, including teen communities. We also publish and share papers with researchers on issues related to young people. For example, we have ongoing relationships with groups like the Aspen Institute and the Humanity Center, and we are a founding sponsor of the Digital Wellness Lab run jointly by Harvard University and Boston Children's Hospital. And because safety and well-being are not just Meta issues, but societal issues, we work with researchers in the field to look more broadly at youth experiences on mobile technology and social media, and how to better support youth as they transition through different stages of life.

We have a long track record of using research and close collaboration with our Safety Advisory Council, Youth Advisors, Suicide and Self-Injury Advisory Group, and additional experts and organizations to inform changes to our apps and provide resources for the people who use them. These relationships and our research efforts have been instrumental in helping develop a number of the tools and features described above, including Take a Break, Quiet Mode, Nudges, Hidden Words, and Restrict, among others.

Question 7. **Why does your company have the age limit of 13 years old for a user to sign up for an account?**

a. Why not younger or older?

We develop our services both to comply with the Children's Online Privacy Protection Act of 1998 (COPPA) and to meet and exceed the high standards of parents and families. We also note that the minimum age requirement of 13 is relatively standard across our industry.

Question 8. **How many minors use your platform? How much money does your company make annually from these minors?**

When we look at the self-reported age of people who are active daily on our apps in the US, about 6% of those who use Instagram daily are teens under 18, and 1% of those who use Facebook daily are teens under 18. However, we do not have full visibility into usage by age group as some people on our apps may register with an inaccurate age. This is among the reasons why we are calling for age verification at the app store level, as teens and parents already provide app store operators with this information when they purchase their devices and set up their accounts.

Question 9. **What percentage of your employees work on trust and safety and how much money does your company invest annually in trust and safety?**

We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone.

Question 10. **It is sometimes challenging for law enforcement conducting criminal investigations to determine the true identity of a person behind a name on social media or other online platforms, and whether an online identity is an actual person. What are you doing to validate the true identity of users – or the fact that a user is a human – when they create an account on your platforms?**

On Facebook, we require people to use their real identities; pretending to be someone else is an explicit violation of our policies. Account compromise is a highly adversarial space across the internet and imposter accounts affect real people—we remove these accounts when we discover them. On Instagram, people are allowed to choose a username, but, as noted above, we comply with valid law enforcement process, including requests for basic subscriber information, such as email and IP addresses. Indeed, law enforcement have indicated that this information is important when conducting investigations.

Like any tech platform, we authenticate people by relying on the information they have added to their accounts to ensure they are who they say they are—like email addresses and phone numbers. To make sure that we only grant access to authentic account owners, we use a combination of automated and manual systems to review these requests off a variety of signals to help us detect potentially suspicious activity and validate legitimate access attempts. In some cases, we may ask for additional information which only the rightful account holder would know in order to restore access and prevent abuse.

Question 11. **Is your company using safety technology to detect and prevent live video child sexual abuse on your platforms and apps that allow users to stream or share live video? If not, please explain.**

a. Has your company tested that or similar technology? If not, are you developing similar technology to address child sexual abuse in live video?

We believe that any instance of content that violates our Child Sexual Exploitation, Abuse, and Nudity policy is one too many, which is why we continue to invest heavily in combating online child exploitation.

Over the years, we have invested heavily in sophisticated technology that helps us proactively find violating content—including in Live—and accounts of this kind and remove them. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in identifying certain content for human review.

Since 2019, we have also made two technologies—PDQ and TMK-PDQF—publicly available that detect identical and nearly identical photos and videos. We use PhotoDNA and other photo- and video-matching technologies that detect identical or near-identical photos and videos of known child exploitative content, and we use Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess. We also use technology to detect and remove Instagram Reels and Stories that violate our Community Guidelines, including by scanning for CSAM terms and for CSE indicators.

In August 2023 alone, we disabled more than half a million accounts on Facebook and Instagram for violating our Child Sexual Exploitation, Abuse, and Nudity policies. And after launching a new automated enforcement effort in September, we saw five times as many automated deletions of Instagram Lives that contained nudity and sexual activity.

The overwhelming majority of people use Facebook and Instagram Live for positive purposes, like sharing a moment with friends or raising awareness for a cause they care about. Still, Live can be abused, and we have taken steps to limit that abuse. When we become aware of CSAM on Live we remove it, report it, and take action against the account responsible.

Question 12. **How are you measuring if your trust and safety policies, practices, and tools are effective in protecting children from sexual abuse and exploitation on your platform?**

a. What specific metrics or key performance indicators do you use?

Over the years, we have invested heavily in sophisticated technology that helps us proactively find violating content and accounts of this kind and remove them. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in identifying certain content for human review.

Our goal is to minimize the impact caused by violations of our policies on people using our services. We measure the viewership prevalence of violating content to gauge how we are performing against that goal. Prevalence estimates how much content that is determined to violate our policies people actually see. Views of content that violates our Child Sexual Exploitation, Abuse, and Nudity policy are infrequent, and we remove much of this content before people even see it. Because of the relative infrequency of violating samples, precisely estimating prevalence is difficult, but we continue making progress in our effort to do so. We estimate the prevalence of child sexual exploitation content on Instagram to be less than 0.01% (or fewer than 100 views in every 1 million). However, we believe that any instance of this content is one too many, which is why we continue to invest heavily in combating online child exploitation.

Question 13. **Is your company using language analysis tools to detect grooming activities? If not, please explain.**

a. What investments will your company make to develop new or improve existing tools?

Using our apps to harm children is abhorrent and unacceptable. Our industry-leading efforts to combat child exploitation focus on preventing abuse, detecting and reporting content that violates our policies, and working with experts and authorities to keep children safe. For years, we have used technology to help find child exploitative content and to help prevent potentially suspicious adult accounts from finding, following, or interacting with young people. As discussed in more detail below, we use language analysis tools to detect potentially suspicious adult accounts, as well as to offer search interventions—or "interstitials"—containing deterrence and prevention messaging.

We work to help keep teens safe by stopping unwanted contact between teens and adults they do not know or do not want to hear from.

We use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts. Specifically, we work to ensure that teens are not recommended to potentially suspicious adult accounts, and potentially suspicious adult accounts are not recommended to anyone (including to teens or other potentially suspicious adult accounts). We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior.

We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations, and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

In addition, when people search for certain terms or hashtags on Instagram that may be related to harmful content, Meta deploys a pop-up interstitial screen following the search to provide deterrence and prevention messaging, and to connect people with expert information and resources. These interstitials are used for child exploitation, among other contexts. For example, people who search for terms associated with suicide and self-injury, eating disorders, or the sale of illicit drugs receive an interstitial directing them to resources to assist in getting help.

Question 14. **What resources have you developed for victims and survivors of abuse on your platforms?**

We have developed more than 50 tools, features and resources to support teens and their parents, including victims and survivors of abuse. As relevant here, these features include both the ability to block accounts, and Restrict, a tool by which people can restrict someone from commenting on their account. Once enabled, comments from a restricted person will only be visible to that person. A person can choose to view the comment, approve the comment so everyone can see it, delete it, or ignore it. We developed Restrict specifically in response to feedback from teens, because they told us they wanted a more subtle way to block bullies without them knowing they had been blocked. In addition, we recognize that our platforms are places where people share deeply personal moments, and from time to time, people may see friends struggling with their mental health and in need of support. That is why we developed anonymous reporting. If a person believes that someone they care about is struggling with their mental health, they can report it anonymously, and we direct those accounts to resources, such as crisis support partners, on Instagram.

Our collaborative work to address child safety does not stop with improving our own services. We also are deeply committed to improving the entire ecosystem and have engaged with child safety nonprofits and academic researchers to complete child safety research with fieldwide impact. Our efforts with these professionals also include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, among other things. Additionally, to help safety stakeholders identify and respond to the most high priority reports at NCMEC, we funded and helped rebuild their case management tool to ensure investigators can get to the most important cases quickly.

We also provide information to people about other programs that help people report their non-consensual intimate images posted online to other participating technology companies, in an effort to aid in preventing the images from being reshared. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support and technical guidance to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

We have also worked with Thorn, a nonprofit that builds technology to defend children from sexual abuse, to develop updated guidance for teens on how to take back control if someone is sextorting them. It also includes advice for parents and teachers on how to support their teens or students if they are affected by these scams. These resources can be found in our updated Sextortion hub within Meta's Safety Center.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Additionally, we collaborate across the industry through organizations like the Technology Coalition, an industry association dedicated solely to eradicating the sexual exploitation of children online. In 2020, Meta joined Google, Microsoft, and 15 other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual abuse. This project includes a renewed commitment and investment from the Technology Coalition,

expanding its scope and impact to protect kids online and help guide its future work. Project Protect focuses on five key areas: tech innovation, collective action, independent research, information and knowledge sharing, and transparency and accountability. We also announced our recent participation in Lantern, a Tech Coalition program that enables technology companies to share a variety of signals about accounts and behaviors that violate their child safety policies. Lantern participants can use this information to conduct investigations on their own platforms and take action. Meta was a founding member of Lantern, providing the Tech Coalition with the technical infrastructure that sits behind the program and encouraging our industry partners to use it. We manage and oversee the technology with the Tech Coalition, ensuring it is simple to use and provides our partners with the information they need to track down potential predators on their own platforms.

Question 15. **What is your response to requests for content removal from CSAM survivors and other members of the public?**

We work to find, remove, and report child sexual abuse material and disrupt the networks of criminals behind it. We promptly disable accounts for various violations of our child exploitation policies, such as the apparent malicious distribution of CSAM or sexual solicitation of children. As required by law, we report all apparent instances of child exploitation identified on our site from anywhere in the world to NCMEC, which coordinates with law enforcement authorities from around the world.

Additionally, we incorporate feedback from survivors in a number of ways, including collaborating with organizations who work with survivors in a safe, trauma-informed way, and meeting survivors at conferences hosted by various stakeholders. Meta provides funding to NCMEC's free service to help survivors and families impacted by online sexual exploitation, and maintains a dedicated reporting channel where NCMEC staff are able to flag non-CSAM content and activity on our platforms that threatens or otherwise identifies CSAM survivors for our review and action.

Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

We have developed ways to help people control their own experience. For example, people can choose who can message them, and can block anyone they do not want to hear from. People can report nude or sexual photos or videos of themselves or threats to share these images or videos to our apps or technologies to prevent them from being reshared. Our teams review reports 24/7 in

more than 70 languages. We have articles in our Help Center that help people understand how to report this activity on Facebook,[13] Instagram,[14] and Messenger.[15]

***Question 16*. While I understand the importance of and support the use of end-to-end encryption (E2EE) to ensure privacy and safety in many online spaces, child protection organizations and law enforcement have raised major concerns about Meta's move to encrypt Messenger. As you make the transition to E2EE on Messenger, how do you plan to address the fact that certain material such as child sexual abuse material (CSAM) that you're currently reporting to NCMEC will become invisible to you? Have you developed an estimate of the anticipated percentage reduction in CSAM detection and reporting to NCMEC?**

End-to-end encryption is a technology that is now widely used by communications services. It is designed to ensure that the contents of a message sent from one device to another can only be understood by the recipient device. It also has important benefits for Meta's users, including helping protect people's privacy and security, by keeping the content of messages confidential between sender and recipient devices.

To address the potential for harm, we have developed a number of tools that can help us proactively detect accounts engaged in potentially malicious patterns of behavior, a capability that helps us spot and address problems at a broad scale across our services. In addition, our machine learning technology can look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect potentially suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors. To help us respond to violations of our policies quickly, we also encourage people to report messages to us in both encrypted and unencrypted services.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

[13] How do I report an abusive photo on Facebook? | Facebook Help Center
[14] How to Report Things | Instagram Help Center
[15] Reporting Conversations | Messenger Help Center (facebook.com)

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

Importantly, implementation of encryption on Messenger does not undercut our commitment to work with law enforcement, nor does it mean we will stop reporting harmful content to the National Center for Missing and Exploited Children (NCMEC). Indeed, we have spent more than a decade developing policies and technologies to help keep young people safe and to keep predators from attempting to use our services to connect with one another. Our comprehensive approach includes cutting-edge technology to prevent, detect, remove, and report violations of our policies that prohibit child exploitation, as well as providing resources and support to victims. We work with professionals, collaborate with industry, and support law enforcement around the world to fight the online exploitation of children. For example, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations.

After the introduction of end-to-end encryption in Messenger, we expect to continue providing more reports to NCMEC than all of our peers, thanks to our industry-leading work on keeping people safe. For example, WhatsApp—which has long been encrypted—removes hundreds of thousands of accounts per month for suspected CSAM violations. In 2022, WhatsApp also made over one million reports to NCMEC, all without breaking encryption. This was significantly more than all other encrypted messaging services combined. As another data point, of the reports we made to NCMEC in the first quarter of 2021 alone, 1.2 million were generated without scanning private messages, which was more than half of the 2.3 million total reports to NCMEC from the rest of industry in all of 2021.

NCMEC has acknowledged that Meta continues to be an industry leader in this work. We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

***Question 17*. Meta has stated that they are introducing default E2EE on Messenger to protect privacy. What is Meta's plan to prioritize the privacy of children and survivors whose CSAM lives on their platforms, exposing the worst moments of their lives to strangers every day?**

We want teens to have safe, age-appropriate experiences on our apps, including on our encrypted services. Encryption helps protect people's privacy and security, by keeping the content of messages confidential between sender and recipient. Moreover, encryption is already widely used by other large messaging services to protect people's private messages and provide people with the privacy and security they expect when messaging friends and family. That is why we will continue to support encryption, while putting features in place to help keep people safe.

As we expand encryption to Messenger and Instagram Direct Messages, our approach to safety is focused on three key elements: (i) preventing potential harm in the first place; (ii) giving people ways to control their experience; and (iii) responding to violations of our policies quickly. This approach is detailed in our whitepaper, Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging, and is further discussed below.

With respect to preventing potential harm in the first instance, as discussed in response to your Question 16, we use machine learning to proactively detect accounts engaged in potentially malicious patterns of behavior. We have built tools and policies specifically to help young people manage interactions with adults. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists. We also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or

added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

To help us respond to violations of our policies quickly, we encourage people to report messages to us in both encrypted and unencrypted services. We have made our reporting tools easier to find and started encouraging teens to report at relevant moments, such as when they block someone. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting potentially suspicious activity and educating people on how to take action. These notices help

people avoid scams, spot impersonations and flag accounts that have been exhibiting potentially suspicious behavior that attempt to connect to minors.

We also continue to educate people to avoid sharing child exploitation content. We know that sometimes people repost sexual images and videos of children in outrage or to raise awareness, and understand that reposting such content, even without malicious intent, re-victimizes the child. In 2021, we launched a video campaign on Facebook called "Report It. Don't Share It." in partnership with child safety organizations to encourage people to stop and think before resharing those images online and to report them to us instead. We also show notices to people to not share these images or videos, directing them to reporting tools. We also offer education and awareness resources related to sextortion, as referenced in our response to your Question 2.

We have also announced efforts to help teens—and their parents and teachers—feel better equipped against those trying to exploit them by distributing intimate images, and supporting creators and safety organizations around the world to address this type of abuse. Specifically, Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We also provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Finally, because child sexual exploitation is an industry-wide concern, we collaborate across the industry through organizations like the Technology Coalition, an industry association dedicated solely to eradicating the sexual exploitation of children online. In 2020, Meta joined Google, Microsoft, and 15 other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual abuse. This project includes a renewed commitment and investment from the Technology Coalition, expanding its scope and impact to protect kids online and guide its work for years to come. Project Protect focuses on five key areas: tech innovation, collective action, independent research, information and knowledge sharing, and transparency and accountability. We also announced our recent participation in Lantern, a program from the Tech Coalition that enables technology companies to share a variety of signals about accounts and behaviors that violate their child safety policies. Lantern participants can use this information to conduct investigations on their own platforms and take action. Meta was a founding member of Lantern, providing the Tech Coalition with the technical infrastructure that sits behind the program and encouraging our industry partners to use it. We manage and oversee the technology with the Tech Coalition, ensuring it is simple to use and provides our partners with the information they need to track down potential predators on their own platforms.

***Question 18*. What investments will your company make in technical solutions to detect CSAM in E2EE environments?**

Please see the responses to your Questions 16 and 17.

***Question 19*. In your testimony you stated that according to a National Academies of Science report, there is no evidence that social media is harmful to children's mental health. Are you aware that the same report suggests there needs to be a comprehensive study conducted on the effects to children and that the report further suggests a national industry standard?**

The consensus study by the National Academies of Science discussed at the hearing notes that the study's "review of published literature did not support the conclusion that social media causes changes in adolescent health at the population level."[16] Instead, the consensus study highlights that "[c]ontrary to the current cultural narrative that social media is universally harmful to adolescents, the reality is more complicated. Social media can connect adolescents with their friends and family and can serve as a place of safety and support . . . and can also serve as an educational resource and help cultivate and expand hobbies, interests, and creative pursuits"[17]

Understanding how technology impacts lives, especially teens' lives, is an important part of what we do. We agree that more research is needed to understand the bigger picture, and we are supporting that research. For example, it is why we supported more funding for research in these areas, like passage of the Children and Media Research Advancement Act, which provides funding to the National Institutes of Health (NIH) to study the impact of technology and media on the development of children and teens. Additionally, we recently announced a pilot program, in partnership with the Center for Open Science, designed to contribute to the public's scientific understanding of how different factors may or may not impact well-being and inform productive conversations about how to help people thrive. We also provided the University of Oxford Internet Institute with access to data as they conducted the largest independent scientific study of a social media platform, which found no evidence linking Facebook adoption and negative well-being.

With so much of our kids' lives spent on mobile devices and social media, it is important to ask and think about any effects on teens—especially on mental health and well-being. Mental health is a complex issue, and the existing body of scientific work has not shown a causal link between using social media and young people having worse mental health outcomes. The National

[16]Consensus Study Report Highlights by National Academies' staff based on the Consensus Study Report Social Media and Adolescent Health (2023), https://nap.nationalacademies.org/resource/27396/Highlights_for_Social_Media_and_Adolescent_Health.pdf.
[17] *Id.*

Academies of Sciences report you reference evaluated results from more than 300 studies and determined that the research "did not support the conclusion that social media causes changes in adolescent mental health at the population level." It also suggested that social media can provide significant positive benefits when young people use it to express themselves, explore, and connect with others. We will continue to monitor research in this area and remain vigilant against any emerging risks.

a. **Do you agree with establishing a national industry standard in regards to online child sexual exploitation? If so, why do you not support the EARN IT Act?**

Please see the response to your Question 1.

***Question 20*. You testified that you "pioneered" a quarterly report for your community standards across different categories of harmful content. Please provide us with copies of those reports for the past five years.**

We have published the Community Standards Enforcement Report since 2018 to more effectively track our progress and demonstrate our continued commitment to making Facebook and Instagram safe and inclusive. Community Standards Enforcement Reports and past data are made available at https://transparency.fb.com/reports/community-standards-enforcement/.

***Question 21*. Meta's approach to help keep people safe when messaging through Messenger or Instagram includes "giving people more controls to help them protect their experience on our apps". Can you explain the logic of how a child under the age of 18 would be able to differentiate or comprehend safe versus nefarious communication on your apps?**

We take the issues of safety and well-being on our apps very seriously, especially for the youngest people who use our services. That is why we apply baseline protections for young people across our apps and default everyone who is under 16 years old in the US into a private account when they join either Instagram or Facebook. On Instagram, a person has a private account, people have to request to follow them to see their posts, Stories, and Reels unless they choose to allow others to reshare their content. People also cannot comment on their content in those places, and they will not see their content at all in places like Explore or hashtags. We also have other defaults in place when people under 18 first sign up for Instagram, including not allowing people they do not follow to tag or mention them, or include their content in Reels Remixes or Guides. And for Facebook, everyone who is under the age of 16 in the US is defaulted into more private settings when they join Facebook, including, restricting:

- Who can see their friends list;
- Who can see the people, Pages and lists they follow;

- Who can see posts they are tagged in on their profile;
- Who is allowed to comment on their public posts; and
- Minors' contact info, school and birthday from appearing in search to a public audience.

These default privacy settings also allow teens to review posts they're tagged in before the post appears on their profile.

Additionally, to help protect teens from unwanted contact, in the US, we have turned off the ability for teens under the age of 16 to receive direct messages from anyone they do not follow or who they are not connected to on Instagram—including other teens—by default. Under this default setting, teens can only be messaged or added to group chats by people they already follow or are connected to, helping teens and their parents feel even more confident that they will not hear from people they do not know in their direct messages. Teens under 16 in supervised accounts will need to get their parent's permission to change this setting. We are also making these changes to teens' default settings on Messenger, where in the US, people under 16 will only receive messages from Facebook friends, or people they are connected to through phone contacts, for example.

Question 22. **Meta ensures that the information provided in reports to NCMEC is actionable by law enforcement – how are you determining what is actionable? Do you have attorneys or legal counsel conducting reviews? Are they current or former prosecutors?**

We work with professionals, collaborate with industry, and support law enforcement around the world to fight the online exploitation of children. Our specialist teams—including former FBI investigators, victim advocates, and federal child safety prosecutors, including those who specialized in CSAM cases—are focused on understanding the patterns and behaviors of people who exploit our platforms so we can continue to adapt and scale our enforcement, and improve our protections.

We are proud of the strong relationship we have developed with NCMEC and continue to report all CSAM found globally to NCMEC's CyberTipline across our family of apps. We have built systems and review processes to prioritize and appropriately action violating content or accounts and, when appropriate, report it to NCMEC or law enforcement.

NCMEC has acknowledged Meta as an industry leader in this work and that Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs." We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

As a general matter, Electronic Service Providers are legally obligated to report apparent violations of laws related to child sexual abuse material they become aware of to NCMEC's CyberTipline. To do so, we submit electronic reports that contain the apparent child exploitative image(s). We endeavor to make our reports robust and include various types of information allowed by law in order to protect people and our services. Additionally, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations. We provide operational guidelines to law enforcement who seek records from Facebook or Instagram.

In addition to reporting content we become aware of, we go beyond the legal requirements and use sophisticated technology to proactively seek out this content, and as a result we find and report more CSAM to NCMEC than any other service today. We make this technology available to the industry to help protect children from exploitation across the internet. For example, we find and report far more content to NCMEC than any other internet service today. In 2022, we made over 26 million reports to NCMEC between Facebook and Instagram. The rest of the industry made less than 6 million reports collectively.

We dedicate significant resources to addressing the concerns of law enforcement authorities and ensuring the timely processing of legal requests. We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone. Meta has a dedicated team to manage law enforcement data requests, including those that involve emergencies and threats to life. Additionally, specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations review potentially violating content and report findings to NCMEC. Our teams are supported by both in-house and retained outside counsel who are experts in laws that apply to government requests for data, and online safety issues.

We will continue collaborating with organizations like NCMEC and child safety experts, while working to prevent the spread of CSAM online.

Questions from Senator Booker

Question 1. **Trust and safety teams are a vital component in combatting the spread of CSAM, hate speech, violence, and other violative content on tech platforms. Despite this, tech companies have time and time again disinvested from their trust and safety team, especially during changes in leadership.**

a. **How has the size of your trust and safety team changed over the past five years? Please provide numbers for each of the past five years.**

Since 2016, Meta has significantly expanded the number of people who work on safety and security. By 2018, Meta doubled the number of people who work on safety issues from 10,000 to 20,000, which includes content reviewers, systems engineers, and security experts. By 2020, Meta built a global team of 35,000 people to work on safety and security. And by 2022, Meta had more than quadrupled the number of people working on safety and security since 2016 to over 40,000 people. We continue to have around 40,000 people devoted to safety and security efforts.

b. **Do your trust and safety teams make submissions to the National Center for Missing & Exploited Children's CyberTipline, or is that a separate unit?**

We have multiple teams that support our efforts with regard to the National Center for Missing and Exploited Children. Their work includes, but is not limited to:

- Building and improving our technology that flags content and activity reportable to NCMEC;
- Building and improving our technology that supports our submissions to NCMEC, as well as reviewing the content that is included in our submissions, as appropriate;
- Building and improving our technology based on new information, such as hashes of newly discovered CSAM that we receive from NCMEC and signals that receive from industry peers through Lantern;
- Conducting investigations into people who use our platforms and potential predatory networks for manual submissions to NCMEC;
- Reviewing and responding to data requests arising from NCMEC reports, as well as preserving relevant data associated with such reports or requests;
- Identifying and supporting projects to help improve NCMEC's systems including its case management tool, and advancing our shared goal of improving child safety, such as through the creation of Take It Down; and

- Communicating with NCMEC—often on a daily basis—and other key stakeholders to deploy an effective, coordinated response to rapidly evolving adversarial behaviors.

We also have a team dedicated to engaging with NCMEC, Child Exploitation and Online Protection Command, Interpol, the FBI, Homeland Security Investigations and numerous other local, federal, and international law enforcement organizations and departments to help them have the information and training needed to make the best use of this process and support efforts to improve process. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC to help safeguard the child.

In addition to reporting content we become aware of, we go beyond legal requirements and use sophisticated technology to proactively seek out this content, and as a result we find and report more CSAM to NCMEC than any other service today. We make this technology available to the industry to help protect children from exploitation across the internet.

While NCMEC already publishes the total number of CyberTips it receives from ESPs on an annual basis, we have started publishing additional data that demonstrates the types of reports we are making to NCMEC. You can access this data in the Meta Transparency Center with this: https://transparency.fb.com/. We will continue collaborating with organizations like NCMEC and child safety experts to protect teens from unwanted contact with adults, while working to prevent the spread of CSAM online.

c. If it is a separate unit, how many members are on the team and how have those numbers changed over the past five years. Please provide numbers for each of the past five years.

Please see the response to your Question 1(b).

Question 2. **The National Center for Missing & Exploited Children's CyberTipline plays an integral role in combatting child sexual exploitation. The tipline helps law enforcement investigate potential cases and allows prosecutors to bring justice to victims. While federal law requires your company to report to the CyberTipline any apparent violations of federal laws prohibiting child sexual abuse material of which you are aware, there are many gaps.**

a. Is there a standard format your reports to the CyberTipline follow? If so, what is that format?

We are proud of the strong relationship we have developed with NCMEC and continue to report all CSAM found globally to NCMEC's CyberTipline across our family of apps. We have built robust systems and review processes to prioritize and appropriately action violating content and accounts and, when appropriate, report it to NCMEC or law enforcement.

As a general matter, Electronic Service Providers are legally obligated to report apparent violations of laws related to child sexual abuse material they become aware of to NCMEC's CyberTipline. To do so, we submit electronic reports that contain the apparent child exploitative image(s). We endeavor to make our reports robust and also include various types of information allowed by law in order to protect people and our services.

In addition to reporting content we become aware of, we go beyond legal requirements and use sophisticated technology to proactively seek out this content, and as a result we find and report more CSAM to NCMEC than any other service today. We make this technology available to the industry to help protect children from exploitation across the internet.

Indeed, NCMEC has acknowledged that Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs." We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts. We will continue collaborating with organizations like NCMEC and child safety experts, while working to prevent the spread of CSAM online.

We support efforts to develop common industry standards on child exploitation, including standards related to Cybertips. In order to do that well it is important to understand that companies across the industry provide a wide variety of services, and as a result of these differences they have access to different types of information to include in these reports. Accordingly, each company's report may vary based on a variety of factors, including information available and accessible to each provider. Additionally, industry standards must balance the feasibility of more detailed robust reporting with the need for timely submissions.

b. Does your company proactively report planned or imminent offenses?

Yes. Prior to 2018, the reporting statute (18 U.S.C § 2258A) did not permit companies to report planned or imminent violations of which they became aware. Meta raised this issue with lawmakers so that when providers have a good faith belief of this type of emergency, they have the means to report it to NCMEC. This provision was carefully crafted as permissive, rather than mandatory, to ensure that highly actionable, emergency reports are sent to NCMEC. A mandatory requirement would likely result in high levels of non-actionable reporting to avoid potential legal risk for failure to report, defeating the goals of the original amendment to the law.

For additional information, please see the response to your Question 2(a).

c. Does your company proactively report potential offenses involving coercion or enticement of children?

Yes, we report potential offenses as enumerated in 18 U.S. Code § 2258A. Please also see the response to your Questions 2(a) and 2(b).

d. Does your company proactively report apparent child sex trafficking?

Yes. Child trafficking is horrific and has no place on our services. We have dedicated teams and invest in sophisticated technology to proactively detect and stop human trafficking. When we become aware of content on Facebook and Instagram that violates our human trafficking policy, we remove it, and, where appropriate, we refer content to relevant authorities, including NCMEC.

We also respond to law enforcement requests related to sex trafficking. We engage with agencies across the world that are dedicated to combating sex trafficking and help inform prevention efforts on our services. We have developed strong relationships with NCMEC, Internet Watch Foundation, ECPAT International, the US Department of Health and Human Services' Office of Child Support Enforcement, and other NGOs to disrupt and prevent sex trafficking online. We also work closely with leading organizations dedicated to fighting trafficking and supporting victims in addition to NCMEC, such as Tech Against Trafficking, Stop the Traffik, and other global NGOs.

Questions from Senator Butler

***Question 1*. Family and parental control tools: I was glad to hear that you have spent time talking with parents and what their families need from your products. I was also glad to hear your companies have a Family Center, or other similar tools, to give parents more insight and control over how their children are using your platforms and apps.**

a. **How do you advertise this feature to parents?**
b. **Can you share data on how many Family Center/parental tools users there are in proportion to total minors on your platforms and products?**

We have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

We are constantly working to make sure parents know about and can choose to use parental control features. We reach parents in a variety of ways, including through Family Center's Education Hub, advertising campaigns, in-app promotion, our work with safety partners, and events with parents. We also collect input from teens in multiple ways. For example, we have hosted programs with organizations such as Girls Scouts of America and the National Parent Teacher Association to create awareness of our safety tools and to get feedback from teens. We have also launched TTC Labs, a global co-design program, that invites young people, parents and experts to participate as collaborators in our design process, empowering them to have their say and ensuring our products meet their needs.

Our ads encouraging parents to use our youth well-being tools and features were seen more than one billion times by tens of million people in the United States since August 2022. Additionally, in 2023, we hosted "Screen Smart" events in six cities (NYC, LA, Miami, Chicago, Nashville, and Seattle), bringing together over 250+ parenting influencers and local stakeholders to educate them on the tools, features, and resources we provide to support parents and teens. Attendees shared content and information from the events about our tools, features, and resources, and those pieces of content reached nearly 50 million impressions. And we recently announced a

series of Screen Smart workshops to help empower parents to confidently manage their teens' usage of smartphones and devices—including on Meta's platforms.

Nonetheless, it can be challenging for parents to supervise the many apps that their teens may use, which is one of the reasons we support federal legislation at the app store level that would make it simpler for parents to oversee their teens' online lives. Parents want to be involved in their teen's online lives, and recent Pew research suggests that 81% of US adults support requiring parental consent for teens to create a social media account. But technology is constantly changing and keeping up with all the apps teens use can feel impossible. As an industry, we should come together with lawmakers to create simple, efficient ways for parents to oversee their teens' online experiences.

That is why we also support federal legislation that would make it simpler for parents to oversee their teens' online lives, including legislation that would require app stores to get parents' approval when teens under 16 download an app. According to a recent Morning Consult poll,[18] parents across both sides of the aisle overwhelmingly support this approach. 81% of Democratic-leaning and 79% of Republican-leaning parents back federal legislation for parental approval of teen app downloads. Over 75% of parents prefer app stores as more secure and straightforward venues for approving downloads, and a more effective method than individual app-level.

In addition to offering a simpler way for parents to approve their teens' app downloads, federal legislation needs to create standards for all apps to adhere to in areas like age-appropriate content, age verification, and parental controls.

We want to help find workable solutions and earlier this year we proposed a framework for legislation.[19] We designed this framework to create clear, consistent standards for all apps, to empower parents and guardians, and to preserve user privacy, in ways that are technologically feasible for the industry. This framework would:

- **Require app stores to get parental approval for teens under 16 to download an app.** Empowering parents to approve their teens' app downloads ensures that they oversee their teens' online experience. Placing the point of approval within the app store simplifies the process and leverages optional approval systems already offered by app stores. App stores would notify parents and request their approval when their teen wants to download an app, including Instagram.

[18] Morning Consult Survey
[19] A framework for legislation to support parents and protect teens online (January 16, 2024)

- **Require certain apps, including social media apps, to offer supervision tools for teens under 16 that parents can activate and control.** Parents should have the tools they need to guide and support their teens online. Certain apps, including social media apps, should be required to offer some form of parental supervision tools, including the ability to set daily time limits on teens' usage, see which accounts their teen is following or friends with, and more. Furthermore, apps can quickly and easily implement these tools if a parent relationship is established in the app store.

- **Require app stores to verify age and provide apps and developers with this information.** Knowing a person's age helps ensure that apps can easily place teens in the right experience for their age group, but parents and teens should not have to provide sensitive information like government IDs to hundreds of apps to verify their age. Parents already provide this information when they purchase a teen's phone and set up their teen's account. App stores have this information and not only can they ease the burden on parents by sharing it with apps, they can help ensure teens are placed in age-appropriate experiences.

- **Require industry to develop consistent age-appropriate content standards across the apps teens use.** Parents are eager to have a better understanding of the content available to their teens and to have guidelines to help them evaluate whether an app is appropriate for their child. We need broader alignment across industry on the types of content companies should consider age appropriate, as there is for other media like movies and video games. It is time we have common industry standards for what is age-appropriate that parents can rely on.

- **Establish national standards to unify the complicated patchwork of inconsistent state laws, and that apply to all apps consistently.** Parents expect consistent standards across all the apps their teens access—regardless of where their teens access or use them.

- **Require industry to develop ads targeting and delivery standards that, for example, limit the personalization of ads for teens under 16 to age and location only.** Industry standards on ad targeting and delivery can help to ensure teens see relevant ads for age-appropriate products and services in their community (e.g., a college prep course) while eliminating the ability to target this audience based on online behaviors or activity. Personalizing ads by age and location is common across industries: for example, advertisers may place relevant ads during teens' TV shows, or in magazines or newspaper sections designed for teens.

Question 2. **Whistleblower allegations about youth mental health: Mr. Zuckerberg, Meta has had two whistleblowers testify at high profile Congressional hearings and talk about**

the harms to young girls perpetuated on your platforms, including the plastic surgery and other "beauty" filters I asked you about. Just a few months ago Mr. Arturo Bejar told us that Meta knew about harms teenagers experience on their platforms. He specifically informed you about these harms.

a. **Mr. Zuckerberg, how do you ensure that employee concerns, especially those related to youth mental health and that of marginalized communities, are taken seriously and addressed in a timely manner?**

We value our employees' opinions, and we take suggestions and proposals related to safety seriously. Meta hires people who care deeply about these issues, and we expect them to ask questions, propose ideas, and challenge leaders of the company. It is how we have made so much progress. Meta has actually implemented many employee proposals, and while we may not adopt every proposal, we take input seriously and make informed decisions about the best way forward.

We have spent more than a decade working on these issues and have developed more than 50 tools, features and resources to support teens and their parents. We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone. We regularly consult with experts in adolescent development, psychology, and mental health to make our platforms safe and age-appropriate for young people, including improving our understanding of which types of content may be less appropriate for young people.

Questions from Senator Coons

Question 1. **While Meta discloses the "prevalence" of content that violates its suicide and self-harm policy, i.e., the percentage of all content on the platform that violates the policy, you testified that the company does not disclose an estimate of the total amount of content on its platforms that violates its suicide and self-harm policy.**

a. **Does Meta measure an estimated total amount of content on its platforms that violates the suicide and self-harm policy? If not, why not?**
b. **If Meta does measure an estimated total amount of content on its platforms that violates the suicide and self-harm policy, why does Meta choose not to disclose this metric?**

Our policies prohibit content on our apps that intentionally or unintentionally celebrates or promotes suicide, self-injury, or eating disorders, and any self-injury content which is graphic, regardless of context. We use a combination of user reports and technology to find this type of policy-violating content, and when we find it, we remove it, regardless of the context or the person's motivation for sharing it. Many times we do not find enough violating samples of suicide and self-injury to precisely estimate prevalence because we remove much of this content before people see it. As a result, we can only estimate an upper limit of how often someone would see content that violates these policies. In the third quarter of 2023, we reported that the upper limit of suicide and self-injury prevalence was about 0.05%, or no more than five views for every 10,000 views on Facebook and Instagram.

Question 2. **Meta does disclose how much content it removes under the platform's suicide and self-harm policy.**

a. **For content that has been removed, does Meta measure how many views that content received prior to being removed? If not, why not?**

Please see the response to your Question 1.

b. **For content that has been removed, does Meta disclose how many views that content received prior to being removed? If so, please provide a specific citation to where Meta discloses that information. If not, why not?**

Our primary metric is prevalence, which tells us how often content that violates our standards is seen relative to the total amount of times any content is seen on Facebook and Instagram. Prevalence considers all the views of content on Facebook or Instagram and measures the estimated percentage of those views that were of violating content. For more information on prevalence, please see the response to your Question 1.

c. Please provide an estimate of the number of views content that was removed under this policy received in January 2024.

Views of violating content that contains suicide and self-injury are very infrequent, and we remove much of this content before people see it. As a result, many times we do not find enough violating samples to precisely estimate prevalence.

In Q4 2023, this was true for violations of our policies on suicide and self-injury, terrorism and restricted goods and services on Facebook and Instagram. In these cases, we can estimate an upper limit of how often someone would see content that violates these policies.

In Q4 2023, the upper limit was 0.05% for violations of our policy for suicide and self-injury on both Facebook and Instagram. This means that out of every 10,000 views of content on Facebook, we estimate no more than 5 of those views contained content that violated the policy.

d. For content that has been removed, does Meta measure demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If not, why not?

We do not track information in the manner requested. For more information, please see the response to your Question 1.

e. For content that has been removed, does Meta disclose demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If so, please provide a specific citation to where Meta discloses that information. If not, why not?

We do not track information in the manner requested. For more information, please see the response to your Question 1.

f. Does Meta measure the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If not, why not?

We do not track information in the manner requested. For more information, please see the response to your Question 1.

g. Does Meta disclose the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If so, please provide a specific citation to where Meta discloses that information. If not, why not?

We do not track information in the manner requested. For more information, please see the response to your Question 1.

Question 3. **Facebook and Instagram utilize algorithms to recommend or amplify content to users.**

a. **For content that has been removed, does Meta measure whether and the extent to which the removed content was recommended or amplified by Meta? If not, why not?**

On Facebook and Instagram, we make recommendations to help people discover new communities and content. We make recommendations based on content people have expressed interest in and actions they take on our apps. We personalize these recommendations with the goal of making them relevant and of value to each individual.

Facebook's Community Standards and Instagram's Community Guidelines govern what content is permissible on our platforms. If we become aware of content that violates these policies, we take steps to remove it. We remove millions of violating posts and accounts every day on both Facebook and Instagram. Most of this happens automatically, with technology working behind the scenes to remove violating content—often before anyone sees it—thereby, minimizing, or in some cases, eliminating distribution. If content is removed for violating our Community Standards, for example, it does not appear in Feed at all.

b. **For content that has been removed, does Meta disclose whether and the extent to which the removed content was recommended or amplified by Meta? If so, please provide a specific citation to where Meta discloses that information. If not, why not?**

Please see response to your Question 3(a).

c. **For content that has been removed, does Meta measure how many views the removed content received after having been recommended or amplified? If not, why not?**

Please see response to your Question 3(a).

d. **For content that has been removed, does Meta disclose the number of views the removed content received after having been amplified or recommended? If so, please provide a specific citation to where Meta discloses that information. If not, why not?**

Please see response to your Question 3(a).

Question 4. **Does Meta support creating industry-wide transparency requirements to disclose basic safety information, like those included in the Platform Accountability and Transparency Act?**

As a general matter, we support the idea of common industry standards for protecting people who use our platforms. At Meta, we believe that more transparency regarding the treatment of content on our platform is a good thing for the public and for the industry. For this reason, we support legislation creating industry-wide transparency requirements to disclose basic safety information, similar to those found in Sec. 10(e) of the Platform Accountability and Transparency Act. We appreciate the productive and collaborative engagement with your staff on the Platform Accountability and Transparency Act, and we look forward to continuing engagement with your office on this bill.

Questions from Senator Cruz

Question 1. **In the last two years, has an employee or commissioner of the Federal Trade Commission (FTC) requested to evaluate or evaluated your data used for training Large Language Models or algorithms or the sources of such data for bias, discrimination, or misinformation?**

Meta routinely briefs the FTC about new products and initiatives. This includes briefings on our Large Language Models and other generative AI products. In connection with those briefings, the FTC has not made requests for or sought details about the matters referenced above.

Question 2. **In the last two years, has an employee or commissioner of the FTC sought details regarding your company's measures related to filtering or blocking inputs and outputs of a Large Language Model or algorithms.**

a. **If yes, has the FTC attempted to coerce or otherwise request you to implement input/output filtering in order to allegedly comply with federal law?**

Meta routinely briefs the FTC about new products and initiatives. This includes briefings on our Large Language Models and other generative AI products. In connection with those briefings, the FTC has not made requests for or sought details about the matters referenced above.

Question 3. **In the last two years, has an employee or commissioner of the Federal Trade Commission sought to evaluate your company's use of measures, including "prebunking" or "debunking", designed to counteract so called "online misinformation"?**

Meta routinely briefs the FTC about new products and initiatives. This includes briefings on our Large Language Models and other generative AI products. In connection with those briefings, the FTC has not made requests for or sought details about the matters referenced above.

Question 4. **In June 2022, the FTC released a report titled "Combatting Online Harms Through Innovation." In this report, the FTC discussed how the deployment of AI tools intended to detect or otherwise address harmful online content is accelerating but may never be appropriate as an alternative to human judgment.**

a. **In the context of protecting children from online harms to what extent does your company rely on automated tools to detect online harm vs. human review? Please be specific.**
b. **What benefits can AI provide to helping detect and/or stop harmful content to children online?**
c. **What does a human reviewer provide that an AI or automated tool cannot? Will we always need some measure of human review in assessing online harms to children?**

d. The FTC has sent mixed signals in its enforcement of COPPA. While the Commission emphasizes not over relying on use of automated tools or AI, they have nonetheless found liability for using human review as alternative signaling overreliance on automated tools. What improvements, if any, should Congress make to clarify the legal tension between use of automated detection tools vs. human review?

People on Facebook and Instagram post billions of pieces of content every day. We have thousands of reviewers around the world. But it is impossible for them to review it all by themselves. That is where Meta artificial intelligence (AI) comes in. We remove millions of violating posts and accounts every day on Facebook and Instagram. Most of this happens automatically, with technology working behind the scenes to remove violating content—often before anyone sees it. Other times, our technology will detect potentially violating content but send it to review teams to check and take action on it.

Over the years, we have invested heavily in sophisticated technology that helps us proactively find violating content and accounts and remove them. For example, AI has improved to the point that it can detect violations across a wide variety of areas without relying on people to report content, often with greater accuracy than reports from humans. This helps us detect harmful content and prevent it from being seen by hundreds or thousands of people. Further, instead of simply looking at reported content in chronological order, our AI prioritizes the most critical content to be reviewed, whether it was reported to us or detected by our proactive systems. This ranking system prioritizes the content that is most harmful based on multiple factors such as virality, severity of harm, and likelihood of violation. In an instance where our systems are near-certain that content is breaking our rules, it may remove it. Where there is less certainty, it will prioritize the content for teams to review. Technology has also helped scale the work of our content reviewers in areas where there may be a higher frequency of violations. By using technology to help in content determinations, our reviewers can focus on determinations where more expertise is needed to understand context and nuance of a particular situation.

AI has helped to advance our content review process and greatly improved our ability to moderate content at scale. But there are still areas where human review is critical. For example, some determinations, such as whether someone is the target of bullying, require an understanding of nuance and context. Human review is helpful in those instances. And our technology relies on training data from reviews done by our teams to identify relevant patterns of behavior and find potentially violating content.

While we use AI technology to help enforce our policies, our use of generative AI tools for this purpose has been limited. But we are optimistic that generative AI could help us take down harmful content faster and more accurately. It could also be useful in enforcing our policies

during moments of heightened risk, like elections. We have started testing Large Language Models (LLMs) by training them on our Community Standards to help determine whether a piece of content violates our policies. These initial tests suggest the LLMs can perform better than existing machine learning models. We are also using LLMs to remove content from review queues in certain circumstances when we are highly confident it does not violate our policies. This frees up capacity for our reviewers to focus on content, which requires more nuance to understand things like tone to determine whether it violates our policies.

Our work to protect people online is never finished. Bad actors will keep trying to evade our technology, so we need to keep improving. That is why our content review system continues to rely on both human review and technology. Our expert teams focus on cases where it is essential to have human review, and we leverage technology to help us scale our efforts in areas where it can be most effective.

Question 5. **In 2021, Congress directed the FTC to research and report on how AI can be used positively to detect and combat fraudulent or deceptive content online. Rather than viewing AI as a potential solution to our online woes, the FTC instead issued a report that read more like an indictment of the technology.**

a. Please explain whether, in your view, AI can be used to positively detect and combat fraudulent or deceptive content, including the recent use of deepfakes or other scams to harm consumers.

Meta has dedicated significant resources to detecting potentially violating content on our services, including detecting AI-generated content such as deepfakes, that violates our policies. Our approach to addressing manipulated media has several components, including working to investigate deceptive behaviors like fake accounts and misleading manipulated media, and our third-party fact-checking program, in which fact checkers rate misinformation, including content that has been edited or synthesized in a way that could mislead people. Our investments have allowed us to build technologies to help proactively identify potentially violating content, prioritize critical content for review, and act on content that violates our policies. We enforce our policies through a combination of people and technology that work to identify violations of our Community Standards across the billions of pieces of content that are posted to our services every day.

Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in enqueuing certain content for human review. For example, our systems flag content that may violate our policies, including fraudulent content and scams, people who use our apps report content to us they believe is questionable, and our own teams review certain content. We work to remove content that violates our policies quickly and at scale. We have also built a parallel content review system to flag posts that may be going viral—no

matter what type of content it is—as an additional safety net. This helps us catch content that our traditional systems may not pick up. We use this tool to detect and review Facebook and Instagram posts that were likely to go viral and take action if that content violated our policies.

Addressing the challenge of deepfakes requires a whole-of-industry approach, which is why we engage with academia, government, and industry. Specifically, we work with industry peers to align on technologies that can make it easier for us and other providers to identify when someone shares content that has been AI-generated. This approach will also pose challenges, as new companies creating AI tools will constantly emerge. Moreover, we know that bad actors will continue trying to find ways to circumvent our detection capabilities. To that end, we continue to partner with the Partnership on AI, in the hope of developing common standards for identifying and labeling AI-generated content, as well as mitigating deceptive AI-generated content, across the industry. In particular, we support efforts to develop industry standards about how and when to apply watermarks to photorealistic images—which is why we welcomed the White House's Voluntary Commitments on AI on this point. We think this is a place where Congress can help drive the consensus forward.

Further, as the difference between human and synthetic content gets blurred, we understand people want to know where the boundary lies. That is why we have been working with industry partners to align on common technical standards that signal when a piece of content has been created using AI. For example, along with twenty other companies in the industry, Meta has pledged to help prevent deceptive AI content from interfering with this year's global elections. The "Tech Accord to Combat Deceptive Use of AI in 2024 Elections" is a set of commitments to deploy technology countering harmful AI-generated content meant to deceive voters. Signatories, including Meta, pledge to work collaboratively on tools to detect and address online distribution of such AI content, drive educational campaigns, and provide transparency, among other concrete steps. Being able to detect these signals will make it possible for us to label certain AI-generated images that people generate or modify with AI off our platforms and post publicly to Facebook and Instagram. We are building this capability now, and in the coming months we will start applying labels in all languages supported by our apps.

In addition to our efforts regarding AI-generated content, we also continue to invest in automated detection technology to improve our ability to detect violating content and help keep people safe. Whether it is improving an existing system or introducing a new one, these investments help us automate determinations on content so we can respond faster and reduce mistakes. The following are some of the technological investments we have made to improve how our tools understand content:

- We developed a new architecture called Linformer, which analyzes content on Facebook and Instagram in different regions around the world.

- We built a new system called Reinforced Integrity Optimizer, which learns from online signals to improve our ability to detect hate speech.
- We improved an image-matching tool called SimSearchNet, which helps our technology detect subtle distinctions in content so we can take action on misinformation.
- We incorporated language tools called XLM and XLM-R, which help us build classifiers that understand the same concept in multiple languages. This means when our technology can learn in one language, it can improve its performance in others, which is particularly useful for languages that are less common on the internet.
- We built a whole entity understanding system, which analyzes content to help determine whether it contains hate speech.

While we use AI technology to help enforce our policies, our use of generative AI tools for this purpose has been limited. But we are optimistic that AI could help us take down harmful content faster and more accurately. It could also be useful in enforcing our policies during moments of heightened risk, like elections. We have started testing Large Language Models (LLMs) by training them on our Community Standards to help determine whether a piece of content violates our policies. These initial tests suggest the LLMs can perform better than existing machine learning models. We are also using LLMs to remove content from review queues in certain circumstances when our systems are highly confident it does not violate our policies. This frees up capacity for our reviewers to focus on content, which requires more nuance to understand things like tone to determine whether it violates our policies.

b. Has the FTC ever consulted with your company to learn how your company deploys AI to better detect and combat fraudulent or deceptive content? Has the DOJ? How about the Federal Elections Commission?

Meta routinely briefs the FTC about new products and initiatives, including briefings on our Large Language Models and other generative AI products.

c. How can Congress empower agencies to use AI positively for the protection of American consumers from fraudulent or deceptive content?

At Meta, we constantly consider how to develop and deploy AI technologies responsibly. We know that AI has brought—and will continue to bring—huge advancements to society, but we also recognize that it comes with risks and the potential to cause unintended consequences. This issue is not unique to Meta, but rather, it is one that stakeholders across the industry must work to address.

For our part, we are working to help advance the responsible design and operations of AI technology and we are committed to building this technology thoughtfully from the start. Progress and responsibility have to go hand in hand. Working together across industry, government, and civil society is essential if we are to develop common standards around safe and trustworthy AI. In particular, we support efforts to develop industry standards about how and when to apply watermarks to photorealistic images—and we think this is a place where Congress can help drive the consensus forward. Until then, we have been working with industry partners to align on common technical standards, which will make it possible for us to label certain AI-generated images that users post to Facebook, Instagram and Threads. We are building this capability now, and in the coming months we will start applying labels in all languages supported by each app. We look forward to working with agencies to share what we have learned from building AI technologies in an open way over the last decade so that the benefits of AI can continue to be shared by everyone.

***Question 6*. Please provide a description of your company's policy regarding the sale or transfer of the data of American users collected on your platform to a third party, including data brokers.**

Meta does not sell people's information to anyone, and we never will. We impose strict restrictions on how our partners can use and disclose the data we provide. Our Privacy Policy makes clear the circumstances in which we work with third-parties that help us provide and improve our services, which makes it possible to operate our companies and provide our services to people around the world.

Our Privacy Policy contains a description of the ways in which we share data with third parties. As the policy explains, user information may be shared with entities outside of the company in a variety of circumstances:

- **Public Information:** Information that users keep "public" can be seen by anyone, on or off our services. Our Privacy Policy provides links with detailed descriptions of how users can learn more about what information is public, and how they can control their visibility on Meta platforms.

- **Apps, Websites, and Third-Party Integrations On or Using Our Services:** As we have explained above, when people on our platforms choose to use third-party apps, websites, or other services that use or are integrated with our services, those third-parties can receive certain information about what people post or share using their products or services. People can also choose to share all their posts with third-party apps, websites, or services.

- **Third-Party Partners:** We work with a variety of partners who help us provide and improve our services. Such partners include:
    - Partners who use our analytics services to help understand how people are engaging with their posts, listings, Pages, videos, and other content on and off Meta's services;
    - Advertisers, which receive reports about the kinds of people seeing their ads and how their ads are performing. However, we do not share information that personally identifies our users (such as their names or email addresses);
    - Measurement partners, which aggregate information to provide analytics and measurement reports to our partners;
    - Vendors and service providers who support our business, such as by providing technical infrastructure services, facilitating payments, etc.;
    - Researchers and academics, who conduct research that advances scholarship and innovation; and
    - Legal requests (in narrowly defined cases).

Question 7. **Has your company ever sold the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the categories of data sold.**

Meta does not sell people's information to anyone, and we never will.

Question 8. **Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the circumstances underlying the basis for such transfer.**

An individual's data may be subject to requests by government agencies (including national security authorities) when they use our services. We have robust policies to scrutinize every government request no matter which government makes the request. These requests must be made in accordance with applicable law and our policies (including Meta's Privacy Policy), and we produce only the information that is narrowly tailored to respond to each request.

Additionally, law enforcement plays a critical role in keeping people safe, and we have a long history of working successfully with them to address a wide variety of threats. We dedicate significant resources to addressing law enforcement concerns, and we carefully review, validate, and respond to the legal requests we receive from them as soon as possible. This includes

prioritizing requests related to emergency situations. And when we see a credible threat on our platform, we do not hesitate to reach out to law enforcement proactively.

Question 9. **Has your company ever sold the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and the categories of data sold.**

Meta does not sell people's data to any entity.

Question 10. **Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and categories of data transferred.**

An individual's data may be subject to requests by government agencies (including national security authorities) when they use our services. We have robust policies to scrutinize every government request no matter which government makes the request. Meta must comply with valid and compulsory legal requests from US government agencies. These requests must be made in accordance with applicable law and our policies (including Meta's Privacy Policy), and we produce only the information that is narrowly tailored to respond to each request.

Additionally, law enforcement plays a critical role in keeping people safe, and we have a long history of working successfully with them to address a wide variety of threats. We dedicate significant resources to addressing law enforcement concerns, and we carefully review, validate, and respond to the legal requests we receive from them as soon as possible. This includes prioritizing requests related to emergency situations. And when we see a credible threat on our platform, we do not hesitate to reach out to law enforcement proactively.

Question 11. **Does your company have a policy to restrict third party use and/or transfer of data collected from users on your platform? Please be specific, including how you enforce such restrictions and whether such restrictions prohibit the sale or transfer of such data to a government agency, including a foreign government agency.**

Third parties are external parties who do business with Meta but are not owned or operated by Meta and typically fall into two major categories: those who provide a service for Meta (like vendors who provide website design support) and those who build their businesses around our platform (like app or API developers). To mitigate privacy risks posed by third parties that receive access to personal information, we developed a dedicated third party oversight and management program, which is responsible for overseeing third party risks and implementing appropriate privacy safeguards.

As part of this third-party oversight and management program, we have created a third party privacy assessment process for service providers to assess and mitigate privacy risk. Our process requires that these service providers are also bound by contracts containing privacy protections. Their risk profile determines how they are monitored, reassessed, and, where appropriate, which enforcement actions to take as a result of violations, including termination of the engagement.

For third party app developers, we have also designed a formal process for enforcing and offboarding third parties who violate their privacy or security obligations. This includes standards and technical mechanisms that support better developer practices across our platform, including:

- **Data Use Checkup (DUC):** Procedures and infrastructure designed to require third party developers to complete an annual Data Use Checkup (DUC), in which developers certify to the purpose and use of each type of personal information that they request or continue to have access to, and that each purpose and use complies with applicable terms and policies. We have introduced new questions and improved logic to strive for greater accuracy in responses and better comprehension from developers. We have also created new tooling to centralize developer communications and requests for additional information into a single location.
- **Monitoring Developer Compliance:** We have developed technical and administrative mechanisms to monitor developers' compliance with our Platform Terms on both an ongoing and periodic basis. When we detect a violation, we take standardized enforcement actions, which, among other factors, take into account the severity, nature and impact of the violation, the developer's malicious conduct or history of violations, and applicable law when determining the appropriate enforcement action to take.
- **Data Security Standards:** We have also developed data security principles based on industry standards for developers to drive better security practices across our platform and the developer ecosystem more broadly.
- **Developer Trust Center:** We launched the Developer Trust Center, a central hub on the Meta for Developers site that brings together material for third party developers on data privacy, data security, Platform Terms, and monitoring mechanisms that they interact with such as App Review, App Re-Review, DUC, and the Data Protection Assessment (DPA).

Question 12. **Between July 4, 2023 and July 14, 2023, was your company contacted by any employee of or contractor for any of the following agencies? Please answer "yes" or "no"**

for each agency and, if "yes," provide the date(s) of contact and the name(s) of the agency employees or contractors that contacted your company.

a. **U.S. Department of Health and Human Services (HHS)**
b. **National Institute of Allergy and Infectious Diseases (NIAID)**
c. **Centers for Disease Control and Prevention (CDC)**
d. **U.S. Food and Drug Administration (FDA)**
e. **The National Institutes of Health (NIH)**
f. **U.S. Department of Homeland Security (DHS)**
g. **DHS Cybersecurity and Infrastructure Security Agency (CISA)**
h. **U.S. Census Bureau**
i. **Federal Bureau of Investigation (FBI)**
j. **U.S. Department of Justice (DOJ)**
k. **The White House Executive Office of the President (EOP)**
l. **U.S. Department of State**

We have received contact for years from individuals across various agencies, including a number of those listed, during Democratic and Republican Administrations. We also consult with experts as we work to provide people with a safe and positive experience on our services. These consultations have included conversations over many years with members of Democratic and Republican Administrations, as well as Democratic and Republican legislative branch officials. We do not share the names of individuals who contact us or who we engage with for a number of reasons—among them safety and security concerns and the fact that those individuals may not want to be named.

We are aware of the preliminary injunction issued by the United States District Court for the Western District of Louisiana on July 4, 2023 that restricts the ability of certain Executive Branch officials to communicate with certain companies. That preliminary injunction was subsequently modified by the United States Court of Appeals for the Fifth Circuit and stayed by the Supreme Court of the United States. The litigation is currently pending before the Supreme Court.

Question 13. **Is it your company's policy to prevent children under 13 from using your social media app(s) or creating an account?**

Meta recognizes the need to keep people who are too young off Facebook and Instagram. Both Facebook's Terms of Service and Instagram's Terms of Use in the United States require people to be at least 13 years old to sign up for Facebook and Instagram. Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the

opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

Children under 13 are only permitted on certain, parent-managed services—specifically Messenger Kids and Meta Quest—if their parent has provided verifiable parental consent for their use of our services and our receipt of their data.

Helping to keep young people safe online is one of our most important responsibilities. Understanding age online is a complex, industry-wide challenge. We have come to understand that people, including young people, sometimes misrepresent how old they are. We believe that the difficulty in understanding someone's age online is not unique to Meta or even to social media, and it warrants a simple solution that would apply across the industry. That is why we support federal legislation that requires app stores to get parents' approval whenever their teens under 16 download apps. With this solution, when a teen wants to download an app—including ours—app stores would be required to notify their parents, much like when parents are notified if their teen attempts to make a purchase. Parents and guardians can decide if they want to approve the download. Parents and guardians can also verify the age of their teen when setting up their phone, negating the need for everyone to verify their age multiple times across multiple apps. Parents want this type of clear system for age verification and parental control over what apps their kids are using.

This industry-wide solution also helps to preserve privacy. By verifying a teen's age on the app store, individual apps—including Meta's—would not be required to collect potentially sensitive identifying information. Apps would only need the age from the app store to confirm that teens are old enough to register for a platform and place them in the right experiences for their age group. Parents and teens will not need to provide hundreds of apps with information like government IDs. Instead they would provide it in just one place, the app store that comes with the device. In many cases, the app store already is collecting this information for its own purposes.

Question 14. **In your view, would it be appropriate for school-aged children to spend time on or access your company's social media app(s) during class?**

We respectfully defer to school administrators who are best positioned to discuss and set the internet access policies for their students. As indicated above, children under the age of 13 are not permitted to have accounts on Facebook or Instagram. More broadly, Meta's mission is to give people the power to build community and bring the world closer together. As part of that mission, we provide services that may be used in educational contexts. Meta's apps and services

are used for a broad range of purposes, including helping people build education communities. For example, many schools operate Pages on Facebook, people may use Facebook Groups to form school-related study groups, and education-related entities have created Messenger bots to help communicate with people. As another example, metaverse technologies have the potential to transform school lessons, bring teachers and students together remotely in shared spaces, enhance vocational training, and create new opportunities for lifelong learnings.

In addition, parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent; schedule breaks for their teens, such as during school or dinner time; see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively).

Question 15. **As a parent, would you be concerned if your child were able to access your company's social media app(s) during class via a school network or device?**

We respectfully defer to school administrators who are best positioned to discuss and set internet access policies for their students. Children under the age of 13 are not permitted to have accounts on Facebook or Instagram. More broadly, Meta's mission is to give people the power to build community and bring the world closer together. As part of that mission, we provide services that may be used in educational contexts. Meta's apps and services are used for a broad range of purposes, including helping people build education communities. For example, many schools operate Pages on Facebook, people may use Facebook Groups to form school-related study groups, and education-related entities have created Messenger bots to help communicate with people. As another example, metaverse technologies have the potential to transform school lessons, bring teachers and students together remotely in shared spaces, enhance vocational training, and create new opportunities for lifelong learnings.

Question 16. **In your view, should elementary and secondary schools block students' access to your company's social media app(s) on school networks and devices?**

We respectfully defer to school administrators who are best positioned to discuss and set internet access policies for their students. Children under the age of 13 are not permitted to have accounts on Facebook or Instagram. More broadly, Meta's mission is to give people the power to build community and bring the world closer together. As part of that mission, we provide services that may be used in educational contexts. Meta's apps and services are used for a broad range of purposes, including helping people build education communities. For example, many schools

operate Pages on Facebook, people may use Facebook Groups to form school-related study groups, and education-related entities have created Messenger bots to help communicate with people. As another example, metaverse technologies have the potential to transform school lessons, bring teachers and students together remotely in shared spaces, enhance vocational training, and create new opportunities for lifelong learnings.

Question 17. **Do you think that school buses equipped with Wi-Fi should allow children to access your company's social media app(s) via a school bus Wi-Fi network during their rides to and from school?**

We respectfully defer to school administrators who are best positioned to discuss and set the internet access policies for their students. More broadly, Meta's mission is to give people the power to build community and bring the world closer together. As part of that mission, we provide services that may be used in educational contexts. Meta's apps and services are used for a broad range of purposes, including helping people build education communities. For example, many schools operate Pages on Facebook, people may use Facebook Groups to form school-related study groups, and education-related entities have created Messenger bots to help communicate with people. As another example, metaverse technologies have the potential to transform school lessons, bring teachers and students together remotely in shared spaces, enhance vocational training, and create new opportunities for lifelong learnings.

Question 18. **As a parent, do you think it is important to supervise your children's internet access?**

Yes. We have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

Among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on

Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians and teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

That is why we also support federal legislation that would make it simpler for parents to oversee their teens' online lives, including legislation that would require app stores to get parents' approval when teens under 16 download an app. According to a recent Morning Consult poll,[20] parents across both sides of the aisle overwhelmingly support this approach; 81% of Democratic-leaning and 79% of Republican-leaning parents back federal legislation for parental approval of teen app downloads. Over 75% of parents prefer app stores as more secure and straightforward venues for approving downloads, and a more effective method than individual app-level.

In addition to offering a simpler way for parents to approve their teens' app downloads, federal legislation needs to create standards for all apps to adhere to in areas like age-appropriate content. We want to help find workable solutions and earlier this year we proposed a framework for legislation.[21] We designed this framework to create clear, consistent standards for all apps, to empower parents and guardians, and to preserve user privacy, in ways that are technologically feasible for the industry. This framework would:

- **Require app stores to get parental approval for teens under 16 to download an app.** Empowering parents to approve their teens' app downloads ensures that they oversee their teens' online experience. Placing the point of approval within the app store simplifies the process and leverages optional approval systems already offered by app stores. App stores would notify parents and request their approval when their teen wants to download an app, including Instagram.
- **Require certain apps, including social media apps, to offer supervision tools for teens under 16 that parents can activate and control.** Parents should have the tools they need to guide and support their teens online. Certain apps, including social media apps, should be required to offer some form of parental supervision tools, including the ability to set daily time limits on teens' usage, see which accounts their teen is following or friends with, and more. Furthermore, apps can quickly and easily implement these tools if a parent relationship is established in the app store.
- **Require app stores to verify age and provide apps and developers with this information.** Knowing a person's age helps ensure that apps can easily place teens in the

[20] Morning Consult Survey
[21] A framework for legislation to support parents and protect teens online (January 16, 2024)

right experience for their age group, but parents and teens should not have to provide sensitive information like government IDs to hundreds of apps to verify their age. Parents already provide this information when they purchase a teen's phone and set up their teen's account. App stores have this information and not only can they ease the burden on parents by sharing it with apps, they can help ensure teens are placed in age-appropriate experiences.

- **Require industry to develop consistent age-appropriate content standards across the apps teens use.** Parents are eager to have a better understanding of the content available to their teens and to have guidelines to help them evaluate whether an app is appropriate for their child. We need broader alignment across industry on the types of content companies should consider age appropriate, as there is for other media like movies and video games. It is time we have common industry standards for what is age-appropriate that parents can rely on.
- **Establish national standards to unify the complicated patchwork of inconsistent state laws, and that apply to all apps consistently.** Parents expect consistent standards across all the apps their teens access—regardless of where their teens access or use them.
- **Require industry to develop ads targeting and delivery standards that, for example, limit the personalization of ads for teens under 16 to age and location only.** Industry standards on ad targeting and delivery can help to ensure teens see relevant ads for age-appropriate products and services in their community (e.g., a college prep course) while eliminating the ability to target this audience based on online behaviors or activity. Personalizing ads by age and location is common across industries: for example, advertisers may place relevant ads during teens' TV shows, or in magazines or newspaper sections designed for teens.

Question 19. **As a parent, would you be concerned if your child's school allowed your child to access the internet on an unsupervised basis, such as on your child's bus ride to and from school via the school bus Wi-Fi?**

We have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to

public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

Question 20. **Do you think Congress should require schools, as a condition of receiving broadband subsidies through the Federal Communications Commission's E-Rate program (which funds broadband for elementary and secondary schools), to block students' access to your company's social media app(s) from school-run networks?**

We respectfully defer to Congress and school administrators on the appropriate conditions to place on the FCC's E-Rate program funding. As indicated above, children under the age of 13 are not permitted to have accounts on Facebook or Instagram. More broadly, Meta's mission is to give people the power to build community and bring the world closer together. As part of that mission, we provide services that may be used in educational contexts. Meta's apps and services are used for a broad range of purposes, including helping people build education communities. For example, many schools operate Pages on Facebook, people may use Facebook Groups to form school-related study groups, and education-related entities have created Messenger bots to help communicate with people. As another example, metaverse technologies have the potential to transform school lessons, bring teachers and students together remotely in shared spaces, enhance vocational training, and create new opportunities for lifelong learnings.

Question 21. **Do you support the bipartisan Eyes on the Board Act of 2023, S. 3074?**

As a general matter, we support the development of a consistent set of rules and controls across a variety of online services.

We also have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, including during the school day; see who their teens follow and who follows their teen.

Question 22. **Have you, your company, or any foundation associated with you or your company, donated or contributed funding, equipment, or services to any of the following organizations in the last ten years (CY 2013 to CY 2023)?**

a. **Education and Libraries Networks Coalition (EdLiNC)**
b. **Open Technology Institute**
c. **Consortium for School Networking (COSN)**
d. **Funds For Learning**
e. **State Educational Technology Directors Association (SETDA)**

f. **Schools, Health, and Libraries Broadband Coalition (SHLB)**
g. **State E-Rate Coordinators' Alliance (SECA)**
h. **EducationSuperHighway**
i. **All4Ed**
j. **Public Knowledge**
k. **Fight for the Future**
l. **Free Press**
m. **Electronic Frontier Foundation**
n. **Benton Foundation or Benton Institute for Broadband & Society**
o. **Electronic Privacy Information Center**

Detailed grant and investment information regarding the Chan Zuckerberg Initiative is publicly available online.[22] Please find more information about Meta's political engagement linked here: https://about.meta.com/facebook-political-engagement/.

***Question 23*. For each such donation or contribution described in the prior question, please detail (1) the type of donation or contribution, such as financial donation, goods or equipment, services, etc.; (2) who made the donation or contribution; (3) the recipient organization; (4) the year the donation or contribution was made; and (5) the total value of that donation or contribution.**

Detailed grant and investment information regarding the Chan Zuckerberg Initiative is publicly available online.[23] Please find more information about Meta's political engagement linked here: https://about.meta.com/facebook-political-engagement/.

***Question 24.* On November 15, 2023, Antigone Davis, Meta's Global Head of Safety, published a blog post that called for federal legislation that requires app stores to get parents' approval when their teen wants to download an app. We've even seen a pretty extensive lobbying and advertising campaign from Meta to drill down this point, including during your testimony.**

Interestingly, this push comes at a time when Congress is debating legislation to help protect the privacy and safety of children, which would put greater responsibility on companies like Meta. The implication, of course, is that someone else except for Meta should bear the burdens of protecting children online.

a. **Do you also believe that Meta should take responsibility for children and teens on your platforms by seeking parental approval prior to use of your products? Why or why not?**

[22] https://chanzuckerberg.com/grants-ventures
[23] https://chanzuckerberg.com/grants-ventures

Helping to keep young people safe online is one of our most important responsibilities. We have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

We know that parents and teens find our tools helpful because they keep using them. For example, among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians and teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

That is why we also support federal legislation that would make it simpler for parents to oversee their teens' online lives, including legislation that would require app stores to get parents' approval when teens under 16 download an app. According to a recent Morning Consult poll,[24] parents across both sides of the aisle overwhelmingly support this approach; 81% of Democratic-leaning and 79% of Republican-leaning parents back federal legislation for parental approval of teen app downloads. Over 75% of parents prefer app stores as more secure and straightforward venues for approving downloads, and a more effective method than individual app-level.

In addition to offering a simpler way for parents to approve their teens' app downloads, federal legislation needs to create standards for all apps to adhere to in areas like age-appropriate content. We want to help find workable solutions and earlier this year we proposed a framework for legislation.[25] We designed this framework to create clear, consistent standards for all apps, to

[24] Morning Consult Survey
[25] A framework for legislation to support parents and protect teens online (January 16, 2024)

empower parents and guardians, and to preserve user privacy, in ways that are technologically feasible for the industry. This framework would:

- **Require app stores to get parental approval for teens under 16 to download an app.** Empowering parents to approve their teens' app downloads ensures that they oversee their teens' online experience. Placing the point of approval within the app store simplifies the process and leverages optional approval systems already offered by app stores. App stores would notify parents and request their approval when their teen wants to download an app, including Instagram.

- **Require certain apps, including social media apps, to offer supervision tools for teens under 16 that parents can activate and control.** Parents should have the tools they need to guide and support their teens online. Certain apps, including social media apps, should be required to offer some form of parental supervision tools, including the ability to set daily time limits on teens' usage, see which accounts their teen is following or friends with, and more. Furthermore, apps can quickly and easily implement these tools if a parent relationship is established in the app store.

- **Require app stores to verify age and provide apps and developers with this information.** Knowing a person's age helps ensure that apps can easily place teens in the right experience for their age group, but parents and teens should not have to provide sensitive information like government IDs to hundreds of apps to verify their age. Parents already provide this information when they purchase a teen's phone and set up their teen's account. App stores have this information and not only can they ease the burden on parents by sharing it with apps, they can help ensure teens are placed in age-appropriate experiences.

- **Require industry to develop consistent age-appropriate content standards across the apps teens use.** Parents are eager to have a better understanding of the content available to their teens and to have guidelines to help them evaluate whether an app is appropriate for their child. We need broader alignment across industry on the types of content companies should consider age appropriate, as there is for other media like movies and video games. It is time we have common industry standards for what is age-appropriate that parents can rely on.

- **Establish national standards to unify the complicated patchwork of inconsistent state laws, and that apply to all apps consistently.** Parents expect consistent standards across all the apps their teens access—regardless of where their teens access or use them.

- **Require industry to develop ads targeting and delivery standards that, for example, limit the personalization of ads for teens under 16 to age and location only.** Industry standards on ad targeting and delivery can help to ensure teens see relevant ads for age-appropriate products and services in their community (e.g., a college prep course)

while eliminating the ability to target this audience based on online behaviors or activity. Personalizing ads by age and location is common across industries: for example, advertisers may place relevant ads during teens' TV shows, or in magazines or newspaper sections designed for teens.

b. Does Meta have a responsibility to remove under-13 accounts from its platforms?

Meta recognizes the need to keep people who are too young off Facebook and Instagram. Both Facebook's Terms of Service and Instagram's Terms of Use in the United States require people to be at least 13 years old to sign up for Facebook and Instagram. Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

Anyone, including individuals who do not have a Facebook or Instagram account, can report to Meta that someone is or appears to be under the age of 13 by filling out a platform-specific online reporting form. Reported accounts are evaluated either by an automated process or through human review, and in some cases both. Our content reviewers are trained to confirm and remove accounts that appear to be used by people who are underage. While manual review is a labor-and time-intensive process, Meta has taken steps to review accounts flagged as potentially underage as quickly as possible after they are reported. Additionally, Meta may learn that someone is underage directly from the individual, if the person attempts to change the date of birth on their account to a date that would make them under 13. In this case, the individual is automatically placed in an "age checkpoint," and we remove the account if the person cannot verify they are over 13.

c. Please provide the number of reports Facebook and Instagram received over the last 5 years regarding the presence of under-13 accounts and the number of those reports that led to a disabled account?

In the last two quarters of 2021, Meta removed more than 4.8 million accounts on Facebook and 1.7 million accounts on Instagram because they were unable to meet our minimum age requirement.

d. Does a Meta employee review every report made regarding the presence of an under-13 account on its platforms?

e. How does Meta verify whether an account that is the subject of an under-13 account is a user who is under 13?

To effectuate Meta's Facebook's Terms of Service and Instagram's Terms of Use in the United States in requiring people to be at least 13 years old to sign up for these platforms, Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates that they are under the age of 13.

And as noted above, Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

Identifying Potentially Underage Accounts

Anyone, including individuals who do not have a Facebook or Instagram account, can report to Meta that someone is or appears to be under the age of 13 by filling out a platform-specific online reporting form. Reported accounts are evaluated either by an automated process or through human review, and in some cases both. Our content reviewers are trained to confirm and remove accounts that appear to be used by people who are underage. While manual review is a labor- and time-intensive process, Meta has taken steps to review accounts flagged as potentially underage as quickly as possible after they are reported. Additionally, Meta may learn that someone is underage directly from the individual, if the person attempts to change the date of birth on their account to a date that would make them under 13. In this case, the individual is automatically placed in an "age checkpoint," and we remove the account if the person cannot verify they are over 13.

Automated Evaluation

An account that has been flagged as potentially underage will first go through an automated process that determines whether the account should be escalated for human review or immediately allowed to continue using the platforms. Where Meta has evidence indicating that the reported individual is over the age of 13, Meta may automatically permit the person who has been flagged as potentially underage to continue using Facebook or Instagram. For example, this can occur when a human reviewer previously evaluated the account for potential underage usage and approved the individual to continue using the platform following the review (pursuant to review guidelines detailed below), the account was previously placed in an age checkpoint and

the person submitted sufficient documentation demonstrating they were at least 13 years old, or the account is so old that it could not reasonably belong to a person under 13. A flagged account will also be permitted to remain on the platform if the account contains no bio or photos, because, as discussed below, a reviewer relies on this data to evaluate whether the account belongs to an underage person.

Manual Review of Potentially Underage Accounts

Flagged accounts that cannot be resolved through the automated processes described above are directed to human reviewers for further evaluation. Meta employs tens of thousands of human reviewers whose duties include reviewing these Facebook and Instagram accounts to manually look for signs that an account has violated the applicable terms of service or content guidelines, including accounts suspected as belonging to people under 13.

All potentially underage accounts that are manually reviewed are evaluated to determine whether they meet our age requirements. For example, reviewers consider the following factors:

- **Account Bio**: Reviewers first evaluate the account's bio for contextual information or self-admission about a person's real age, including a written statement of the person's age, birth date, or grade in school. Reviewers are also trained to look for additional signals that indicate whether the account holder is underage. An account that contains information that explicitly states or contextually implies that the individual is under 13 will be checkpointed and the person will be required to provide Meta with proof of age.
- **Account Photos**: If the account bio does not contain sufficient written information to determine age, the reviewer will evaluate the photos contained in the account. If a human reviewer finds sufficient signals that the account holder may be under 13, or is unsure of whether an account holder is under 13 based on a review of the account media, the account will immediately be placed in an age checkpoint.

Responding to Potentially Underage Accounts

When Meta identifies a potentially underage account, their account will be placed in an age checkpoint. While in the checkpoint, a person does not have access to their account, and they are shown a blocking screen if they attempt to log into their account. This means checkpointed accounts cannot view or interact with any content or ads on the platform. Additionally, checkpointed accounts are not visible to other people on the platform, and people cannot see or interact with the checkpointed account or the photos or videos posted to it.

If the person is unable to demonstrate that they are 13 years of age or older, their account is permanently disabled and removed, and the data is deleted consistent with Meta's standard deletion policies.

Other Mechanisms for Identifying Potentially Underage Accounts

In addition, we have also partnered with Yoti, a company that offers privacy-preserving ways to verify age. Yoti is verified by the Age Check Certification Scheme and is the leading age verification provider for several industries around the world including social media, gaming and age-restricted e-commerce. Expert and governmental organizations in youth and privacy have publicly endorsed Yoti for their approach and expertise in responsible artificial intelligence.

For example, on both Instagram and Facebook, a person who attempts to change their date of birth to go from the age of under 18 to 18 or older is required to verify their age through one of two options, ID upload or video selfie provided by the third-party Yoti. If Yoti estimates that the person is under the age of 13, the account will be placed in an age checkpoint. As explained above, if the person is unable to demonstrate that they are 13 years of age or older, their account is permanently disabled and removed, and when the account is disabled, the data is deleted consistent with Meta's standard deletion policies.

f. Under what circumstances, does Meta choose to not act on a report made about an under-13 account?

Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual. For more information on how we identify and remove people we believe are underage, please see the response to your Question 24(b).

g. The Children's Online Privacy Protection Act (COPPA) currently requires companies to obtain parental consent for the collection of a child's data if the company has "actual knowledge" or is "directed to children". Is "actual knowledge" a high legal bar for Meta to meet?

We develop and operate our services to comply with the requirements of the Children's Online Privacy Protection Act of 1998 (COPPA). Meta's Terms of Service in the United States require

people to be at least 13 years old to sign up for Facebook and Instagram. As noted above, when Meta identifies a potentially underage account, their account will be placed in an age checkpoint. While in the checkpoint, a person does not have access to their account, and they are shown a blocking screen if they attempt to log into their account. This means checkpointed accounts cannot view or interact with any content or ads on the platform. Additionally, checkpointed accounts are not visible to other people on the platform, and people cannot see or interact with the checkpointed account or the photos or videos posted to it.

If the person is unable to demonstrate that they are 13 years of age or older, their account is permanently disabled and removed, and the data is deleted consistent with Meta's standard deletion policies.

h. In your view, is a company's (such as Meta's) policy to remain "willfully ignorant" of a child user's age meeting the responsibility that you believe you owe children and parents to protect underage users from using your platform?

We respectfully disagree with any such characterization of Meta's approach to determining age online. In the US, we require people to be at least 13 years old to sign up for Instagram or Facebook. In some countries, our minimum age is higher. Working to help keep young people safe online is one of our most important responsibilities, and that is why we have invested over $20 billion in safety and security across our platforms since 2016. And, as noted above, Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

Identifying Potentially Underage Accounts

Anyone, including individuals who do not have a Facebook or Instagram account, can report to Meta that someone is or appears to be under the age of 13 by filling out a platform-specific online reporting form. Reported accounts are evaluated either by an automated process or through human review, and in some cases both. Our content reviewers are trained to confirm and remove accounts that appear to be used by people who are underage. While manual review is a labor-and time-intensive process, Meta has taken steps to review accounts flagged as potentially underage as quickly as possible after they are reported. Additionally, Meta may learn that someone is underage directly from the individual, if the person attempts to change the date of birth on their account to a date that would make them under 13. In this case, the individual is

automatically placed in an "age checkpoint," and we remove the account if the person cannot verify they are over 13.

Automated Evaluation

An account that has been flagged as potentially underage will first go through an automated process that determines whether the account should be escalated for human review or immediately allowed to continue using the platforms. Where Meta has evidence indicating that the reported individual is over the age of 13, Meta may automatically permit the person who has been flagged as potentially underage to continue using Facebook or Instagram. For example, this can occur when a human reviewer previously evaluated the account for potential underage usage and approved the individual to continue using the platform following the review (pursuant to review guidelines detailed below), the account was previously placed in an age checkpoint and the person submitted sufficient documentation demonstrating they were at least 13 years old, or the account is so old that it could not reasonably belong to a person under 13. A flagged account will also be permitted to remain on the platform if the account contains no bio or photos, because, as discussed below, a reviewer relies on this data to evaluate whether the account belongs to an underage person.

Manual Review of Potentially Underage Accounts

Flagged accounts that cannot be resolved through the automated processes described above are directed to human reviewers for further evaluation. Meta employs tens of thousands of human reviewers whose duties include reviewing these Facebook and Instagram accounts to manually look for signs that an account has violated the applicable terms of service or content guidelines, including accounts suspected as belonging to people under 13.

All potentially underage accounts that are manually reviewed are evaluated to determine whether they meet our age requirements. For example, reviewers consider the following factors:

- **Account Bio**: Reviewers first evaluate the account's bio for contextual information or self-admission about a person's real age, including a written statement of the person's age, birth date, or grade in school. Reviewers are also trained to look for additional signals that indicate whether the account holder is underage. An account that contains information that explicitly states or contextually implies that the individual is under 13 will be checkpointed and the person will be required to provide Meta with proof of age.
- **Account Photos**: If the account bio does not contain sufficient written information to determine age, the reviewer will evaluate the photos contained in the account. If a human reviewer finds sufficient signals that the account holder may be under 13, or is unsure of

whether an account holder is under 13 based on a review of the account media, the account will immediately be placed in an age checkpoint.

Responding to Potentially Underage Accounts

When Meta identifies a potentially underage account, their account will be placed in an age checkpoint. While in the checkpoint, a person does not have access to their account, and they are shown a blocking screen if they attempt to log into their account. This means checkpointed accounts cannot view or interact with any content or ads on the platform. Additionally, checkpointed accounts are not visible to other people on the platform, and people cannot see or interact with the checkpointed account or the photos or videos posted to it.

If the person is unable to demonstrate that they are 13 years of age or older, their account is permanently disabled and removed, and the data is deleted consistent with Meta's standard deletion policies.

In addition, we have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

Among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians and teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

That is why we also support federal legislation that would make it simpler for parents to oversee their teens' online lives, including legislation that would require app stores to get parents' approval when teens under 16 download an app. According to a recent Morning Consult poll,[26] parents across both sides of the aisle overwhelmingly support this approach; 81% of Democratic-leaning and 79% of Republican-leaning parents back federal legislation for parental approval of teen app downloads. Over 75% of parents prefer app stores as more secure and straightforward venues for approving downloads, and a more effective method than individual app-level.

In addition to offering a simpler way for parents to approve their teens' app downloads, federal legislation needs to create standards for all apps to adhere to in areas like age-appropriate content. We want to help find workable solutions and earlier this year we proposed a framework for legislation.[27] We designed this framework to create clear, consistent standards for all apps, to empower parents and guardians, and to preserve user privacy, in ways that are technologically feasible for the industry. This framework would:

- **Require app stores to get parental approval for teens under 16 to download an app.** Empowering parents to approve their teens' app downloads ensures that they oversee their teens' online experience. Placing the point of approval within the app store simplifies the process and leverages optional approval systems already offered by app stores. App stores would notify parents and request their approval when their teen wants to download an app, including Instagram.

- **Require certain apps, including social media apps, to offer supervision tools for teens under 16 that parents can activate and control.** Parents should have the tools they need to guide and support their teens online. Certain apps, including social media apps, should be required to offer some form of parental supervision tools, including the ability to set daily time limits on teens' usage, see which accounts their teen is following or friends with, and more. Furthermore, apps can quickly and easily implement these tools if a parent relationship is established in the app store.

- **Require app stores to verify age and provide apps and developers with this information.** Knowing a person's age helps ensure that apps can easily place teens in the right experience for their age group, but parents and teens should not have to provide sensitive information like government IDs to hundreds of apps to verify their age. Parents already provide this information when they purchase a teen's phone and set up their teen's account. App stores have this information and not only can they ease the burden on parents by sharing it with apps, they can help ensure teens are placed in age-appropriate experiences.

[26] Morning Consult Survey
[27] A framework for legislation to support parents and protect teens online (January 16, 2024)

- **Require industry to develop consistent age-appropriate content standards across the apps teens use.** Parents are eager to have a better understanding of the content available to their teens and to have guidelines to help them evaluate whether an app is appropriate for their child. We need broader alignment across industry on the types of content companies should consider age appropriate, as there is for other media like movies and video games. It is time we have common industry standards for what is age-appropriate that parents can rely on.

- **Establish national standards to unify the complicated patchwork of inconsistent state laws, and that apply to all apps consistently.** Parents expect consistent standards across all the apps their teens access—regardless of where their teens access or use them.

- **Require industry to develop ads targeting and delivery standards that, for example, limit the personalization of ads for teens under 16 to age and location only.** Industry standards on ad targeting and delivery can help to ensure teens see relevant ads for age-appropriate products and services in their community (e.g., a college prep course) while eliminating the ability to target this audience based on online behaviors or activity. Personalizing ads by age and location is common across industries: for example, advertisers may place relevant ads during teens' TV shows, or in magazines or newspaper sections designed for teens.

i. Would Meta support increasing the requirements of COPPA's knowledge standard to impose greater responsibility on companies (including Meta) to protect against underage use of a platform? Why or Why Not?

Understanding people's age on the internet is a complex challenge across our industry. This is because some may misrepresent how old they are online, and current methods for verifying age are imperfect and come with privacy and equity costs. Instead of increasing the requirements of COPPA's knowledge standard, we support federal legislation that requires app stores to get parents' approval whenever their teens under 16 download apps. With this solution, when a teen wants to download an app, app stores would be required to notify their parents, much like when parents are notified if their teen attempts to make a purchase. Parents and guardians can decide if they want to approve the download. Parents and guardians can also verify the age of their teen when setting up their phone, negating the need for everyone to verify their age multiple times across multiple apps.

Question 25. **Yes or no: Did employees of or contractors for the Cybersecurity and Infrastructure Security Agency (CISA) ever ask Meta to meet with employees of or contractors for the Department of Homeland Security Office of Inspector General (DHS OIG)?**

a. If yes, provide the date of the request from CISA, the channel through which the request was made, and the name of the CISA employee(s) or contractor(s) who made the request.

We have received contact for years from individuals across various agencies, including CISA, during Democratic and Republican Administrations. We also consult with experts as we work to provide people with a safe and positive experience on our services. These consultations have included conversations over many years with members of Democratic and Republican Administrations, as well as Democratic and Republican legislative branch officials. We do not share the names of individuals who contact us or who we engage with for a number of reasons—among them safety and security concerns and the fact that those individuals may not want to be named.

Question 26. **It has been reported that you have invested $100 million in a technology-based "personalized learning" platform called Summit Learning, claiming it would improve educational outcomes.**

a. What is the current status of Summit Learning?

After reaching over 100 schools and over 20,000 students across the country, the team at Meta (then known as Facebook) stopped working on Summit Learning in 2017, after which the Chan Zuckerberg Initiative became Summit Learning's technology partner. Meta does not have any access to data from the Summit Learning Platform or the broader Program. As a philanthropic philanthropic organization and a separate entity from Meta, Meta cannot speak to the current status of Summit Learning. For more information about the Chan Zuckerberg Initiative, please visit czi.org.

b. What lessons did the pandemic teach you regarding screen-based learning and the use of technology in education?

Digital technologies have transformed education over the last two decades. But there are limits to 2D technologies. While remote learning tools kept the wheels of education turning during the pandemic, anyone with teenage kids can attest to the fact that it was often a frustrating experience. It was hard to keep them engaged for lengthy periods interacting with a flat screen. They lacked that vital sense of presence—interacting with their classmates and teachers in a shared space. The metaverse is the next evolution of the internet—and it is this sense of presence that sets it apart. It spans a range of technologies, including virtual reality headsets that transport you to whole new environments; augmented reality glasses that will one day project computer-generated images onto the world around you; and mixed reality experiences that blend physical and virtual environments.

For most of us, learning is social—we learn from and with others, and from each other's experiences. It is about interaction and discussion as much as it is about absorbing facts. Academic studies have found that VR can positively improve comprehension, knowledge retention, student engagement, attention span and motivation, which is something we all intuitively understand. It is so much easier to remember doing something than being told something. That is what makes the possibilities for learning in the metaverse so exciting. Instead of telling students what the dinosaurs were like, they can walk among them. Entire science laboratories can be built and filled with equipment that most schools would never be able to afford. Medical students can practice complex surgery without risk to patients or themselves.

This is not science fiction or wishful thinking—it is happening right now. A college student in Ohio could attend a seminar led by a professor in Seoul. Students in the most remote corner of Alaska could tour NASA, the Louvre in Paris, or the Grand Egyptian Museum in Cairo. A personal tutor could run a session with a student in a completely different city without either having to leave their house.

Question 27. **How many pieces of content from the Israel-Hamas War have been removed automatically by your systems (i.e., without any human review)?**

a. **For the content [removed automatically], provide a breakdown of the reasons for the content's removal.**
b. **How many of the removals [...] were appealed?**
c. **How many of the appeals [...] have been reviewed?**
d. **How many of the appeals [...] have been granted?**
e. **For the content [removed automatically], do you plan to conduct a policy review of the content to ensure that content in the public interest was not erroneously removed from your platform(s)?**

In the wake of the attack on October 7, 2023, Meta took immediate crisis response measures. As conflict-related content exponentially surged on our platforms, we implemented a number of temporary measures across both Arabic and Hebrew markets, seeking equitable outcomes in limiting the prevalence of violating content on our platforms. We quickly established a special operations center staffed with experts, including fluent Hebrew and Arabic speakers, to closely monitor and respond to this rapidly evolving situation in real time. This allows us to remove content that violates our Community Standards or Community Guidelines faster.

In the nine days following October 7, we removed or marked as disturbing more than 2,200,000 pieces of content in Hebrew and Arabic for violating our policies around DOI, violent and graphic content, hate speech, violence and incitement, bullying and harassment, and coordinating harm. As compared to the two months prior, in the three days following October 7, we removed seven times as many pieces of content on a daily basis for violating our DOI policy in Hebrew and Arabic alone. In the majority of cases, we remove the content before people even see it.

Our internal appeals process allows for erroneous content takedowns to be reversed. These appeal mechanisms are available on both Facebook and Instagram. Both human review teams and technology play a role in reviewing user reports and appeals, and we aim to prioritize appeals with potentially harmful content first.

As an example, stemming from user appeals related to the Israel-Hamas War, on December 7, 2023, Meta's Oversight Board announced it had selected two cases for expedited review, a process by which the Board issues accelerated content decisions within 30 days in exceptional circumstances. On December 19, 2023, the Oversight Board overturned Meta's original decision to remove the content in both cases, finding that restoration of the content to the platform, with a warning screen, is consistent with Meta's content policies, values, and human rights responsibilities. The Oversight Board's guidance in these cases, along with feedback from other experts, will help us to continue to evolve our policies and response to the ongoing Israel-Hamas War.

Question 28. **Describe how international laws requiring certain content moderation, such as the European Union's Digital Services Act, have affected your decisions about what content from the Israel-Hamas War to allow or remove from your platform(s).**

a. **What specific rules and regulations have required you to take down or moderate more content than you may have otherwise if it were not for these laws?**
b. **How would your decisions to remove content pursuant to international laws differ if you faced a legal obligation in the United States to not remove content protected by the First Amendment?**

Meta operates globally and faces legal obligations in markets in which we operate, including the Digital Services Act (DSA) in the EU. We note that the DSA is an EU-specific regulatory scheme that operates within a particular legal and regulatory architecture and applies to the services provided by our EU entities.

Many of the DSA's requirements largely align with our approach to moderating content on our platform, where we set standards (policies) informed by several factors and put in place enforcement actions for content that we find that violates our policies. For example, our content policies for Facebook and Instagram apply to all users and are set out clearly including via our Terms of Service and in the Transparency Center, where we also provide information on how we enforce those policies. Meta removes content worldwide that violates those policies when we become aware of it.

In addition, we have put in place EU-specific measures in keeping with the DSA's requirements. For example, we have reporting mechanisms in place for individuals and entities within the EU to report content to Meta that may be illegal in the EU and/or in an EU member state. This

allows us to take action to restrict content in countries where it may be illegal, even if that content is not uploaded or shared from the same country. We review reports about potentially illegal content carefully, including with regard to our commitments as a member of the Global Network Initiative and our Corporate Human Rights Policy. In some cases, content is reported to us by NGOs or government agencies and is prioritized accordingly.

Question 29. **According to the Wall Street Journal, Meta reported that it had "blocked thousands of hashtags that sexualize children, some with millions of posts, and restricted its systems from recommending users search for terms known to be associated with sex abuse" after receiving queries from the Wall Street Journal.**

a. **Provide the full list of hashtags blocked. Include the number of unique posts and accounts associated with each hashtag.**
b. **Provide the full list of search terms blocked. Include the number of times Meta's systems recommended each search term and note if any search terms were fully blocked (rather than only pulled from recommendations).**
c. **Did Meta receive any user reports regarding any posts or accounts associated with a blocked hashtag? If yes, include the number of reported posts and accounts, the number of reports filed with respect to each post or account, the user-selected reason for each report, whether a human moderator reviewed each report, and what action Meta took (if any) as a result of each report.**

Child exploitation is a horrific crime that we work to fight aggressively on and off our platforms. We have spent more than a decade developing policies and technology to help keep young people safe and to keep predators from attempting to use our service to connect with one another. Our comprehensive approach includes policies prohibiting child exploitation; cutting-edge technology to prevent, detect, remove, and report policy violations; and the provision of resources and support to victims.

We rely on both automated technology, reports, and investigations to take action on violating hashtags, account names, search terms, and emojis. We work to prevent and detect both inherently violating terms and terms that are not inherently violating but may be used by adversarial actors seeking or offering inappropriate content. We work to avoid showing search results for these terms to help prevent the discovery of potentially harmful content. Because we recognize this is a constantly evolving area, we also work with our specialist child safety teams and child safety professionals to help us understand evolutions in coded language and to identify new and evolving terms, phrases, slang, and emojis that could be used in an attempt to evade our detection systems and bypass our policies. Our teams use these signals and technology to proactively find new trends, misspellings and spelling variations of this language, as well as terms and phrases related to child exploitation, that we can input into our systems to proactively find and disrupt efforts to evade our protections.

Last year, we created a task force to address allegations about the effectiveness of our work in this area. As part of that work we reviewed existing policies; examined technology and enforcement systems we have in place; and made changes that strengthened our protections for young people, banned predators, and removed networks they use to connect with one another. Our child safety teams continue to work on additional measures.

Our lists of search terms and hashtags are also not static, but continue to evolve with input from our technology and experts throughout the industry. For example, we have used technology to find relationships between terms that we already know could be harmful or that break our rules and other terms used at the same time. These could be terms searched for in the same session as violating terms, or other hashtags used in a caption that contains a violating hashtag. We combined our systems so that as new terms are added to our central list, they will be actioned across Facebook and Instagram simultaneously. We may send Instagram accounts, Facebook Groups, Pages and Profiles to content reviewers, restrict these terms from producing results in Facebook and Instagram Search, and block hashtags that include these terms on Facebook and Instagram.

As a general matter, we do not share specific lists of blocked hashtags and search terms or provide detailed descriptions of how our tools work or our enforcement efforts, which, if revealed, could provide a roadmap to highly-motivated bad actors who seek to evade our detection and filing of NCMEC reports, which would ultimately undermine our efforts. That said, we are working hard to further augment the measures we have in place so that as predatory behaviors and coded language evolve, so do we.

Regarding your question on user reporting, we encourage user reporting, as it is important to provide helpful context to take action against people who violate our policies and an opportunity to support victims. On Instagram and Facebook, we enable people to report content or conduct they believe violates our policies and flag for our review. We have built systems and review processes to prioritize and appropriately address violating content or accounts, and, when appropriate, report it to NCMEC or law enforcement.

We use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts. Specifically, we work to ensure that teens are not recommended to potentially suspicious adult accounts, and potentially suspicious adult accounts are not recommended to anyone (including to teens or other potentially suspicious adult accounts). We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially

suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Additionally, we have built sophisticated technology that has enabled us to find, remove, and report more exploitative content than any other company that reports to NCMEC. And we will continue to refine our systems, and we call upon the rest of the industry to do the same.

***Question 30*. How many Instagram profiles of users believed to be under the age of 18 were accessed via a hashtag or search term [described above]?**

We do not track information in the manner requested. However, we have numerous policies and tools in place to avoid showing teens' content or accounts in search surfaces on our services.

For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. If a person has a private account, people have to follow the account to see their posts, Stories and Reels. People also cannot comment on others' content in those places, and they will not see this content at all in places like Explore or hashtags. And, as we announced in late 2021, we also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

For teens who choose to have a public account or who opt out of the default settings, we still implement measures to protect them from interacting with potentially suspicious adult accounts. For example, we do not show young people's accounts in Explore, Reels or 'Accounts Suggested For You' to these adults. If they find young people's accounts by searching for their usernames, they are not able to follow or message them (or vice versa). These potentially suspicious adult accounts also are not able to see comments from young people on other people's posts, nor are they able to leave comments on young people's posts. We also prompt teens to review and restrict their privacy settings. When someone comments on a teen's post, tags/mentions them in another post, or includes their content in Reels Remixes or Guides, the teen will receive a notification to review their privacy settings, and will have the option to stop people from interacting with them.

***Question 31*. The Wall Street Journal reported that Meta "permitted users to search for terms that its own algorithms know may be associated with illegal material" and, in those cases, surfaced an interstitial screen warning users that the results might contain "images of child sexual abuse." The interstitial screen […] allowed users to select between "Get resources" or "See results anyway."**

a. **List all search terms that received this interstitial screen warning. Provide an explanation for how these search terms were selected.**
b. **For each search term that received this interstitial screen warning, include the number of unique times the term was searched within the past 365 days and the number of times that the user selected "See results anyway".**
c. **For each search term that received this interstitial screen warning, did Meta's systems ever recommend the term to a user within the past 365 days?**
d. **According to the Wall Street Journal, Meta has since removed the option for users to "See results anyway" when they search for such terms. Describe Meta's initial rationale for providing an option to "See results anyway".**

We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, and content that sexualizes children. We go beyond legal requirements and use sophisticated technology to find, remove, and report child sexual abuse material and disrupt the networks of criminals behind it. We disable accounts that appear to be involved in malicious distribution of CSAM or sexual solicitation of children, and report apparent instances of child exploitation identified on our site to NCMEC, which coordinates with law enforcement authorities from around the world. We report more CSAM to NCMEC than any other service today.

When people search for certain terms or hashtags on Instagram that may be related to harmful content, Meta deploys a pop-up interstitial screen following the search to provide deterrence and prevention messaging, and to connect people with expert information and resources. These interstitials are used for child exploitation, among other contexts. For example, people who search for terms associated with suicide and self-injury, eating disorders, or the sale of illicit drugs receive an interstitial directing them to resources to assist in getting help.

We display interstitial screens in response to searches for terms that relate to violations of our policies, as well as terms that do not relate to violations of our policies, that may potentially be used by people to seek out or offer inappropriate content. Because we employ interstitial screens for non-violating terms, we also include the ability for people to click through to see otherwise non-violating content that these searches return. For example, advocacy groups or other individuals may use terms and hashtags that would otherwise be violating to spread awareness about an issue in a non-violating way.

Regarding your question about the child safety interstitial, we want to make clear that while the "see results anyway" option was available, it was not intended to take people to content that violated our Facebook Community Standards or Instagram Community Guidelines, but was intended instead to allow access to non-violating content. Our technology proactively seeks out

violating content, and we remove it when we find it. Accordingly, any results visible were not expected or intended to contain violating content.

Last year, we removed the ability to click through to see results related to the child safety interstitial on Facebook and Instagram to eliminate any ambiguity. Today, when people search for terms associated with child exploitation or sexual activity, they receive a pop-up blocking screen making it clear that this kind of content is illegal and Meta offers to connect them with help.

Question 32. **Describe in detail the resources Meta has allocated to its child safety teams over the past five years, including numbers of staff disaggregated by policy, investigators, engineering and technical teams, and content reviewers. How do these numbers compare with Meta's teams covering other content areas, such as misinformation?**

Meta has invested more than $20 billion in safety and security across our platforms since 2016, and $5 billion in 2023 alone. Since 2016, Meta has significantly expanded the number of people who work on safety and security. By 2018, Meta doubled the number of people who work on safety issues from 10,000 to 20,000, which includes content reviewers, systems engineers and security experts. By 2020, Meta built a global team of 35,000 people to work on safety and security, including specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to review potentially violating content and report findings to NCMEC. And by 2022, Meta had more than quadrupled the number of people working on safety and security since 2016 to over 40,000 people.

Question 33. **Indicate whether Meta employees had any contact, acting in their capacity as employees of Meta, with officials from the following agencies and departments from January 1, 2016 to present. Please answer "yes" or "no" for each.**

a. **National Security Agency (NSA)**
b. **Central Intelligence Agency (CIA)**
c. **FBI – National Election Command Post**
d. **FBI – Office of Private Sector (OPS) program**
e. **DHS – Office of Intelligence & Analysis (I&A)**
f. **FBI and DHS – Domestic Security Alliance Council (DSAC)**

We have received contact for years from individuals across various agencies, including a number of those listed, during Democratic and Republican Administrations. We also consult with experts as we work to provide people with a safe and positive experience on our services. These consultations have included conversations over many years with members of Democratic and Republican Administrations, as well as Democratic and Republican legislative branch officials. We do not share the names of individuals who contact us or who we engage with for a number of

reasons—among them safety and security concerns and the fact that those individuals may not want to be named.

In general, we have a long history of working successfully with the DOJ, the FBI, DHS, the Defense Department, the State Department, national security officials, and other agencies to address a wide variety of threats. As we have stated before, we work with entities like the FBI's Foreign Influence Task Force to combat malign foreign influence threats on our services, and we recognize the need to work together—across industry and between industry and government—to be successful.

Question 34. **Indicate whether Meta employees, acting in their capacity as employees of Meta, have or have had any contact with the following government officials, acting in their official capacities, from January 1, 2016 to present. Please answer "yes" or "no" for each individual.**

a. **Luke Beckman, DHS CISA**
b. **Wayne Brady, FBI FITF**
c. **Gretchen Burrier, FBI**
d. **William Castle, DOD OGC**
e. **Judy Chock, FBI FITF**
f. **William Cone, FBI FITF**
g. **John Dragseth, DHS CISA CFITF**
h. **Caitlin Durkovich, NSC**
i. **Jen Easterly, DHS CISA**
j. **Luke Giannini, FBI FITF**
k. **Geoff Hale, DHS CISA**
l. **Adam Hickey, DOJ**
m. **Jason Humbert, FDA**
n. **Chris Inglis, National Cyber Director**
o. **Chad Josiah, DHS CISA CFITF**
p. **Chris Krebs, DHS CISA**
q. **Matthew Masterson, DHS CISA**
r. **Sean Newell, DOJ**
s. **Brady Olson, FBI FITF**
t. **Lisa Page, DOJ OGC**
u. **Rodney Patton, DOJ**
v. **Matthew Perry, FBI OGC**
w. **Shelby Pierson, ODNI**
x. **Michael Pollice, FBI New York / DSAC Coordinator**
y. **Lauren Protentis, DHS CISA CFITF**
z. **Michael Purtill, NCTC**

aa. Sean Ragan, FBI San Francisco
bb. Kris Rose, DHS CISA
cc. Robert Schaul, DHS CFITF
dd. Rob Silvers, DHS
ee. Allison Snell, DHS CISA CFITF
ff. John Stafford, DHS CISA CFITF
gg. Johnny Starrunner, FBI OPS
hh. Samaradun Kay Stewart, DOS GEC
ii. Peter Strzok, FBI
jj. Kathryn Tillman, FBI San Francisco
kk. Bryan Vorndran, FBI Assistant Director for Cyber
ll. Fred Whitney, FBI San Francisco / DSAC Coordinator
mm. Carter Wilkinson, FBI OGC
nn. Kim Wyman, DHS CISA CFITF

We have received contact for years from individuals across various agencies, including a number of those listed, during Democratic and Republican Administrations. We also consult with experts as we work to provide people with a safe and positive experience on our services. These consultations have included conversations over many years with members of Democratic and Republican Administrations, as well as Democratic and Republican legislative branch officials. We do not share the names of individuals who contact us or who we engage with for a number of reasons—among them safety and security concerns and the fact that those individuals may not want to be named.

In general, we have a long history of working successfully with the DOJ, the FBI, DHS, the Defense Department, the State Department, national security officials, and other agencies to address a wide variety of threats. As we have stated before, we work with entities like the FBI's Foreign Influence Task Force to combat malign foreign influence threats on our services, and we recognize the need to work together—across industry and between industry and government—to be successful.

Question 35. **Did Meta ever send, regardless of whether solicited, a list of user accounts to an employee of any agency or department listed [above]? If yes, please note the channel of communication and a description of the contents of such list(s), including whether they contained (a) accounts of U.S. citizens and (b) accounts of any U.S. federal, state, or local elected officials.**

We have a long history of working successfully with the DOJ, the FBI, state and local law enforcement, and other government agencies to address a wide variety of threats. We have been able to provide support to authorities around the world. We reach out to law enforcement when

we see a credible threat of imminent offline harm, contacting federal, state, or local law enforcement depending on the specific circumstances of a threat. We also have robust processes in place to handle government requests we receive, and we disclose account records in accordance with our terms of service and applicable law. We have law enforcement response teams available around the clock to respond to emergency requests.

Question 36. **Did Meta ever receive, regardless of whether solicited, requests from or via an employee of any agency or department listed [above] to review, monitor, investigate, promote, or restrict content or accounts related to the following topics? Answer "yes" or "no" for each topic, indicate the requesting agency or department, and describe any actions taken by Meta subsequent to the request.**

a. **Foreign mis- or disinformation, and/or foreign malign influence, related to the 2016, 2018, 2020, and 2022 federal election cycles.**
b. **Voting mis- or disinformation related to the 2016, 2018, 2020, and 2022 federal election cycles.**
c. **The treatment of authoritative information related to voting during the 2016, 2018, 2020, and 2022 federal election cycles.**
d. **Mis- or disinformation related to the COVID-19 pandemic.**
e. **The treatment of authoritative information related to the COVID-19 pandemic.**
f. **Civil unrest related to abortion policy in the United States.**
g. **Civil unrest related to policing practices in the United States.**
h. **The dissemination or publication of any materials from the hard drive of Hunter Biden's laptop.**

We work closely with law enforcement, regulators, election officials, researchers, academics, and civil society groups, among others, to strengthen our services against election interference and the spread of misinformation. This engagement is incredibly important. Sharing information between tech companies, governments, and law enforcement has proven critical to identifying and disrupting foreign interference campaigns early, ahead of elections. As an example, prior to the 2020 elections, we investigated and took down three covert influence operations from Russia, Mexico, and Iran targeting the US, after receiving a tip from US law enforcement about off-platform activity by these threat actors.

As described more below, we have continued to strengthen our internal capacity to detect and enforce against malicious activity since 2017 and continue to engage in threat sharing with experts across our industry and civil society. With respect to our election protection work, we have engaged with state attorneys general and other federal, state, and local law enforcement officials responsible for election protection. When they identified potential voter interference or other violations of our policies, we investigated and took action if warranted, and we have established strong channels of communication to respond to any election-related threats.

When it comes to disinformation, we tackle it through our policies and enforcements against coordinated inauthentic behavior (CIB), which covers coordinated networks that centrally rely on fake accounts to mislead people about who they are and what they are doing to manipulate or corrupt public debate for a strategic goal. We conduct our own independent investigations and enforce against CIB. We do so based on the deceptive behavior we see on our platform, not based on the content they share. Our team focused on disrupting influence operations includes experts across the company, with backgrounds in law enforcement, national security, investigative journalism, cybersecurity, law, internet freedom, human rights, and engineering.

Our technical teams continue to build scaled solutions to help detect and prevent these violating behaviors, and we work with civil society organizations, researchers, and governments to strengthen our defenses. We have also improved our detection systems to more effectively identify and block fake accounts, which are the source of a lot of inauthentic activity.
We regularly publish Adversarial Threat Reports, which detail the results of our efforts to combat CIB, as well as other adversarial threats we detect and remove from our platforms. Our Q4 2023 report can be found at https://transparency.fb.com/metasecurity/threat-reporting. We also report on our integrity enforcement progress publicly in our quarterly Community Standards Enforcement Report. This report includes metrics on how Meta is performing in preventing and removing content that violates our Community Standards and fake accounts.

Regarding COVID-19, we partnered with government agencies throughout the pandemic to connect people to authoritative health information and helpful resources, and we were transparent about the fact that we did so. In developing the standard for imminent physical harm as it relates to COVID-19, we consulted the CDC and other governmental health experts to assess whether a false claim, if believed by an individual, would increase the likelihood that the individual would contract or spread the virus. We updated the claims that we removed based on guidance from health authorities. For other false claims related to COVID-19, we have leveraged our third-party fact-checking program to reduce the distribution of false and misleading content. For example, in May 2021, Facebook stopped removing claims that COVID-19 was man-made, in response to a change in rating from third-party fact checkers.

Importantly, Meta's COVID-19 misinformation policies evolved alongside scientific research throughout the pandemic, and we stopped removing claims that the CDC and other health experts informed us were no longer harmful. We also reassessed whether our policies should remain in place altogether as the threat of COVID-19 subsided, vaccines became more available, and scientific research regarding the pandemic improved. For example, in July 2022, Meta asked its Oversight Board for advice on whether our measures to address dangerous COVID-19 misinformation, introduced in extraordinary circumstances at the onset of the pandemic, should remain in place. The Board advised that we should stop removing those claims in countries that

were no longer experiencing a state of emergency from COVID-19. Based on the Board's advice, we now take a more tailored approach to our COVID-19 misinformation rules consistent with the Board's guidance and our existing policies—our COVID-19 misinformation rules are no longer in effect globally, as the global public health emergency declaration that triggered those rules has been lifted, and we only enforce those specific policies in the few countries still having a COVID-19 public health emergency declaration in place, which the United States does not. We have also narrowed the claims enforced in those countries to only those that are prevalent on our platforms.

Regarding content about the October 14, 2020 *New York Post* story, given what happened in the 2016 election, we were concerned about potential election interference in the 2020 election. To be clear, at no point did we take any action to block or remove the content from our services. This reporting was always available on our services and people could, and did, engage with it. However, given the concerns raised, we took steps to slow the spread of content and provide fact-checkers the opportunity to assess it. After seven days, we lifted the temporary demotion on this content because it was not rated false by an independent fact-checker.

Question 37. **Copies of any unclassified documents, such as memos, threat assessments, joint advisories, or Liaison Information Reports (LIRs), that were provided to Meta by an employee of any agency or department listed [above].**

We consult with a number of external experts and partners as we work to provide people with a safe and positive experience on our services. This can include members of the government, as well as other members of the technology industry, nonprofits, law enforcement, civil society organizations, academics with relevant experience, and more. We do not share the names of all of the groups and individuals we consult with or the information they provide for a number of reasons—among them safety and security concerns and the fact that groups or individuals may not want to be named. We would refer you to the agencies you have specified for additional information.

Question 38. **Provide a complete list of the names of any individuals outside of your organization that you consulted with in developing any of the documents and information [that describe your recommendation systems and any content moderation policies for such systems].**

Academics, experts, and other stakeholders share information with Meta and give feedback on how we might better tackle our policies. We are constantly evaluating—and, where necessary, changing—our content policies. That said, we apply our own policies, and our enforcement decisions might differ from decisions others might make, including those with whom we partner. We do not share the names of all of the groups and individuals we consult with for a number of

reasons—among them safety and security concerns and the fact that groups may not want to be named.

Our Community Standards, published online at https://transparency.fb.com/policies/community-standards/, outline what is and is not allowed on Meta services. We base our policies on principles of voice, safety, dignity, authenticity, and privacy. We also publish our quarterly Community Standards Enforcement Report to give visibility into how we are doing at enforcing the Community Standards. Google and Twitter have their own content moderation policies and make content moderation decisions based on those policies.

***Question 39*. On average, how much additional distribution can a poster expect from being included in your recommendations? Please include a brief summary of your methodology for estimating this percentage.**

Our ranking and recommendations systems are dynamic and highly complex and depend, in part, on the ever-changing interests across the billions of people who use our services. Across our apps, we make personalized recommendations to the people who use our services to help them discover new communities and content. Both Facebook and Instagram may recommend content, accounts, and entities that people do not already follow. Our goal is to make recommendations that are relevant and of value to each person who sees them. For this reason, our recommendations are unique and personalized. Ultimately, we make recommendations based on content people have expressed interest in and actions they take on our apps.

Distribution for any given piece of content included in recommendations can vary widely based on a number of factors, including the type of content, subject matter and how people may have interacted with the content in the past. For example, some content receives reduced distribution in accordance with our Content Distribution Guidelines. Similar to how our Community Standards indicate the types of content that we do not allow on Facebook, our Content Distribution Guidelines describe the types of content we think may either be problematic or low quality, so we reduce its distribution in Feed for everyone. Reduced distribution itself may also vary depending on a number of factors, including the number of times the poster or commenter has violated our rules previously, the degree of confidence of our systems' predictions, among other things. Given the extent of variables affecting distribution, it is extremely challenging to measure the exact distribution recommendations any given piece of content may receive with any degree of precision.

***Question 40*. What percentage of total time spent on your platforms is driven by your recommendation systems? Of that time, what is the median amount of time that users**

spend within a 24-hour period? Please include a brief summary of your methodology for calculating this percentage.

When we make recommendations, our goal is to make them relevant and of value to each person who sees them. In-feed recommendations are increasingly contributing to engagement, and we have seen Reels time become more incremental to overall engagement on our services as we continue to improve our recommendations.

AI-driven feed recommendations continue to grow their impact on incremental engagement. To take one example of such engagement, we previously announced that in Q3 2023, we saw a 7% increase in time spent on Facebook and a 6% increase on Instagram as a result of recommendation improvements. Relatedly, in Q1 2023, AI recommendations drove a more than 24% increase in time spent on Instagram.

Along with surfacing content from friends and family, in Q1 2023, more than 20% of content in Facebook and Instagram feeds were recommended from people, groups, or accounts that people did not follow.

Question 41. **What percentage of total time spent by users under 18 on your platforms is driven by your recommendation systems? Of that time, what is the median amount of time that users under 18 spend within a 24-hour period? Please include a brief summary of your methodology for calculating this percentage.**

When we make recommendations, our goal is to make them relevant and of value to each person who sees them. When it comes to time spent, we want to give people on our platforms—especially teens—tools and resources to help them manage their experiences in the ways that they want and need, including the time they spend. For example, we have a number of features that let teens work with their parents to set daily limits for the total time that teens can spend on our apps, Take A Break notifications, which show full-screen reminders to leave the Instagram app, Quiet Mode, which turns off notifications at night, and Nudges, which include alerts that notify teens that it might be time to look at something different if they have been scrolling on the same topic for a while. Many of these tools—particularly for teens—result in decreased time spent on our platforms.

Question 42. **For the recommendations [viewed by users under 18], please list the top 25 topics, using your internal classifications, associated with the recommended content, entities, or accounts.**

Our recommendations are highly dynamic and personalized, so there is no stable set of topics that comprise a "top 25" list that receives the most distribution. Our Widely Viewed Content

Report on Facebook aims to provide more transparency and context about what people are seeing by sharing the top 20 most-viewed domains, links, Pages, and posts for a given quarter on Feed in the United States. However, it does not break down views by age bracket for a given piece of content. The report provides insights into the various content types that appear on Feed to help people better understand our distribution systems and how they influence the content people see on Facebook.

Our report for Q3 2023 shares data on views and viewers of content in Feed, including recommended content, seen in the United States between July 1, 2023 and September 30, 2023. In Q3 2023, 65% of views came from posts shared by people's friends, from Groups people had joined, or Pages they had followed (see breakdown below). Of the remaining 35% of Feed content views in the US during Q3 2023, 23.7% came from in-Feed recommendations, which show people content from sources they are not connected to, but we think they might be interested in; we refer to this content as "unconnected posts." The last 11.3% came from less common services, such as Events, and logging discrepancies.

In Q3 2023, the following top 20 domains collectively accounted for about 0.6% of all Feed content views in the US during Q3 2023.

1. Youtube.com (123M Content Viewers)
2. Tiktok.com (111.9M Content Viewers)
3. Media1.tenor.co (107.9M Content Viewers)
4. Gofundme.com (105.3M Content Viewers)
5. CBSNews.com (99.3M Content Viewers)
6. People.com (98.7M Content Viewers)
7. Today.com (84.8M Content Viewers)
8. Dailymail.co.uk (83.6M Content Viewers)
9. TMZ.com (80.9M Content Viewers)
10. Rollingstone.com (80.4M Content Viewers)
11. CNN.com (80.3M Content Viewers)
12. ETonline.com (77.2M Content Viewers)
13. Variety.com (71.5M Content Viewers)
14. CBSSports.com (70.9M Content Viewers)
15. Amazon.com (70.7M Content Viewers)
16. NYPost.com (69.2M Content Viewers)
17. Pagesix.com (69.2M Content Viewers)
18. Goodmorningamerica.com (66.7M Content Viewers)
19. SI.com (66.2M Content Viewers)
20. APnews.com (66M Content Viewers)

In Q3 2023, the following top 20 Facebook Pages collectively accounted for about 0.9% of all US content views during Q3 2023.

1. LADbible (140.4M Content Viewers)
2. UNILAD (123.5M Content Viewers)
3. Lessons Learned In Life (113.7M Content Viewers)
4. All You Can Eat (108.7M Content Viewers)
5. People (106.3M Content Viewers)
6. E! News (106.1M Content Viewers)
7. ESPN (105.4M Content Viewers)
8. Sassy (99.4M Content Viewers)
9. MetDaan Tips (96M Content Viewers)
10. BadassAuto (95.6M Content Viewers)
11. SportsCenter (93M Content Viewers)
12. Liz & Jeff (92.7M Content Viewers)
13. Entertainment Tonight (91.8M Content Viewers)
14. Tyla (91.4M Content Viewers)
15. Miranda Blankenship (90.2M Content Viewers)
16. Bleacher Report (90.1M Content Viewers)
17. ABC7 (88M Content Viewers)
18. Daily Mail (87.2M Content Viewers)
19. Lacey's Tattoos, Bikes, Music, and More (87.1M Content Viewers)
20. Simple ideas (85.9M Content Viewers)

Question 43. **For the recommendations [viewed by users under 18], please list the top 100 sources of recommendations.**

Our recommendations are highly dynamic and personalized, so there is no stable set of sources that comprise a "top 100" list that receives the most distribution.

Delivering great content recommendations is an important part of what makes Facebook and Instagram valuable for people around the globe. Our systems show people the relevant content from their particular connections—the friends, accounts, Groups, and Pages they have chosen to follow. But we also deliver highly personalized recommendations from the tens of billions of pieces of content that are outside of a person's network of Facebook or Instagram connections. AI-driven recommendations help people dive deeper into their interests and discover new things while also supporting creators in finding new audiences for their work. As Mark Zuckerberg noted on a recent earnings call, more than 20% of content in a person's Facebook and Instagram feeds is now recommended by AI from a variety of sources, including people, Groups, or accounts they do not follow. A person might regularly like posts about mountain biking, for

example, or perhaps they are part of a community that shares popular biking trails. Our recommendations could show them a post about a unique biking trail from a Page or a Group that they do not happen to follow. Or we could suggest stories about record-breaking races or creators' Reels of their cycling adventures. We could also show something else—such as easy trail mix recipes—that other mountain biking fans found valuable, with the hypothesis that this person, too, might find it interesting. This demonstrates how sources of recommendations can vary widely, including based on people's preferences, interests, and how they interact with content.

Question 44. **Do you place any limits on the total amount of content, accounts, or entities that users can be served by your recommendation systems in a given period of time? If yes, please elaborate. If no, please explain why not.**

The recommended content and accounts that a person sees on our apps is dynamic and based on a number of factors, including the time a person chooses to spend on our apps and the choices they make while using the app. For example, a new user who does not belong to Groups or have any Friends will see a high percentage of recommendations.

We have released a number of system cards for Facebook and Instagram to date, including one on AI systems that recommend "unconnected" content from people, Groups, or accounts they do not follow. The system card on Feed recommendations provides that the AI system behind Facebook Feed Recommendations automatically determines which content shows up in your Feed, and in what order, by predicting what you are most likely to be interested in or engage with. These predictions are based on a variety of factors, including what and whom you have followed, liked or engaged with recently.

People can also limit the amount of recommended content they see, as we provide tools to help them manage their experiences on our platforms however they see fit. For example, on Facebook, people can filter through and browse their Feeds tab, giving them more control over what they see. We also have Favorites, a tool where people can control and prioritize posts from the friends and Pages they care about most in their Feeds tab. Specifically, people can select up to 30 friends and Pages to include in Favorites, and posts from these selections will appear higher in ranked Feed and can also be viewed as a separate filter. In addition, people can select "Show More" or "Show Less" on certain posts from the people and communities they are connected to and posts that Facebook recommends to them. Selecting "Show More" will temporarily increase the ranking score for that post and posts like it. Selecting "Show Less" temporarily decreases the post's ranking score. If people want to temporarily stop seeing suggested content on Facebook, they can "snooze" it for 30 days.

And, on Instagram, we allow people to switch to different feed views, allowing people the option

to see posts in chronological order or a dedicated feed of just the accounts they add to their Favorites. People can also use the "Not Interested" control on Instagram to remove posts from their feed that they do not find interesting or relevant, and we will suggest fewer posts like it in the future. In addition, people can "snooze" all suggested posts in their feed for 30 days. And the Sensitive Content Control on Instagram allows people to choose how much or how little sensitive content they see. People can also make a close friends list on Stories and share with just the people they have added. Our hope is that these tools give people more control over the content they see on our platforms.

Question 45. **Have you ever, or do you currently, maintain any hardcoded lists of individual accounts, entities, or individual pieces of content that are (a) whitelisted or (b) blacklisted from appearing in your recommendation systems? If yes, please provide a description of each list and the number of items on each list.**

Not all content allowed on our platforms, or accounts that post content, will be eligible for recommendation. Through our Recommendations Guidelines, we work to avoid making recommendations that could be low-quality, objectionable, or particularly sensitive, and we also avoid making recommendations that may be inappropriate for younger viewers.

Specifically, our Recommendations Guidelines outline five categories of content that are allowed on our platforms, but that may not be eligible for recommendations.

1. Content that impedes our ability to foster a safe community, such as content that discusses self-harm, suicide, or eating disorders, as well as content that depicts or trivialises themes around death or depression.
2. Sensitive or low-quality content about Health or Finance, such as content that promotes or depicts cosmetic procedures.
3. Content that people broadly tell us they dislike, such as content that includes clickbait.
4. Content that is associated with low-quality publishing, such as unoriginal content that is largely repurposed from another source without adding material value.
5. False or misleading content, such as content including claims that have been found false by independent fact-checkers or certain expert organizations.

Our Recommendations Guidelines also consider whether certain accounts or entities are eligible for recommendation. We try to not recommend accounts (including Profiles and Page admins) or entities (such as Pages, Groups, or Events) that:

1. Recently violated Facebook's Community Standards or Instagram's Community Guidelines.

2. Repeatedly and/or recently shared content (including the names or cover photos associated with groups or Pages) we try not to recommend.
3. Repeatedly posted vaccine-related misinformation that has been widely debunked by leading global health organizations.
4. Repeatedly engaged in misleading practices to build followings, such as purchasing "likes."
5. Have been banned from running ads on our platforms.
6. Recently and repeatedly posted false information as determined by independent third party fact-checkers or certain expert organizations.
7. Are associated with offline movements or organizations that are tied to violence.
8. Discuss or depict suicide and self-harm in the account name, username, profile photo or bio (with the exception of accounts focused on providing support, raising awareness, and recovery).

In addition, people have told us they want to see less political content, so we have spent the last few years refining our approach on Facebook to reduce the amount of political content—including from politicians' accounts—people see in Feed, Reels, Watch, Groups You Should Join, and Pages You May Like. We have recently extended this approach in Reels, Explore and In-Feed Recommendations on Instagram and Threads, too.

As part of this, we aim to avoid making recommendations that could be about politics or political issues, in line with our approach of not recommending certain types of content to those who do not wish to see it.

At the same time, we are preserving people's ability to find and interact with political content that is meaningful to them if that's what they are interested in on Facebook Feed. When ranking political content in Facebook Feed, our AI systems consider personalized signals, like survey responses, that help us understand what is informative, meaningful, or worth people's time. We also consider how likely people are to provide us with negative feedback on posts about political issues when they appear in Facebook Feed. We have shifted away from ranking political content in Facebook Feed based on engagement signals—such as how likely people are to comment on or share content—since we've found that they are not reliable indicators that the content is valuable to someone.

In addition, people can personalize what they see on Facebook through customization tools we offer in Feed Preferences and directly in places like Feed. People can provide direct

feedback on a post by selecting Show more or Show less and use Reduce to adjust the degree to which we demote some content. If someone does not want Meta to personalize their Feed at all, they can use the Feeds tab, which will rank posts chronologically. They can also add people to their Favorites list so they always see content from their favorite accounts.

Question 46. **Have you ever, or do you currently, maintain any hardcoded lists of individual accounts, entities, or individual pieces of content that are (a) boosted or (b) downranked in your recommendation systems? If yes, please provide a description of each list and the number of items on each list.**

Our Content Distribution Guidelines outline types of content that receive reduced distribution in Feed. Our efforts to reduce problematic content in Feed are rooted in our commitment to responding to people's direct feedback, incentivizing creators to invest in high-quality and accurate content, and fostering a safe community—and we continue to adjust and develop our guidelines in line with these values.

Reduced distribution may also depend on the context. During critical moments such as elections, or in situations with elevated risk of violence or other severe human rights risks, we are especially mindful of the need to carefully tailor our approach to keeping people safe while protecting their ability to express themselves. As such, our teams closely monitor trends on our platforms and investigate situations to determine whether and how best to respond. As appropriate, we may apply limited, proportionate, and time-bound measures that can be quickly implemented to address a specific, emerging risk. In some cases, we may further reduce the visibility of certain types of content, above our standard reductions, that may not necessarily violate our Community Standards, but come close to the line. To respond to other risks, we may reduce the distribution of content more significantly if it is posted from accounts that have recently and repeatedly posted violating content, or if it is likely to violate our policies but we need additional time to review. In some circumstances, we may also reduce the distribution of widely shared content in order to slow its overall spread. This is particularly helpful when the content could be misinformation or incite violence. If our teams determine that a piece of content violates our policies, we will remove it, even if its visibility has already been reduced.

There are also circumstances where we may boost certain accounts, entities, or individual pieces of content through recommendations. For example, we may boost authoritative sources in recommendations in order to help ensure people using our services see highly reliable, helpful information during moments of high user need.

Question 47. **Have you ever, or do you currently, include any human-curated content, accounts, or entities in your recommendations? If yes, please describe and provide copies of any curation guidelines.**

Recommendations for accounts that a person may want to follow or like are based on a variety of signals, including the accounts that the person already follows and likes, and the other people that follow and like those accounts. These suggestions are generated using machine learning systems. As a general matter, employees do not determine the rankings or recommendations for any specific piece of content.

One of our primary goals with recommendations is to make it easier for people using our services to discover entertaining and timely content on our apps. As such, we sometimes combine human input with machine learning to help us achieve that. As one example, on Facebook, we have featured timely content related to specific events in culture and entertainment, such as the World Cup, the Super Bowl, and the Grammys.

Question 48. **Please list all U.S.-based users with more than 500,000 total followers or subscribers that have been removed from recommendations, even if temporarily, for a period of at least three continuous days within the past ten years. Please include the duration of and reason for the removal, and note whether the removal is currently in effect.**

Our Recommendations Guidelines provide information on when accounts, pages, or Groups may be removed from recommendations on our platforms. On Facebook, accounts, pages, or Groups that repeatedly violate our policies may be removed from recommendations and have their distribution reduced. In 2021, we launched "Account Status" on Facebook, a feature to help every user understand the penalties Facebook applied to their accounts. It provides information about the penalties on a person's account (currently active penalties as well as past penalties), including the rationale for the penalty. In general, if people have a restriction on their account, they can see their history of certain violations, warnings, and restrictions their account might have, as well as how long this information will stay in Account Status on Facebook.

On Instagram, recently or repeatedly posting policy-violating content can result in your entire account becoming ineligible to be recommended. If your content is removed for violating Community Guidelines, you will be notified in the Instagram app. Professional accounts (business accounts or creator accounts) have the ability to check if their account's content is ineligible to be recommended to non-followers in Account Status. Instagram's Account Status feature can be used to see if content the user posted or something on the user's profile goes against the Recommendations Guidelines.

If an account's content is not eligible to be recommended, the account owner can see a sample of content or components of the profile that may violate the Recommendations Guidelines and any content that has been removed for violating Meta's Community Standards. Users can also edit or

delete posts that may violate the Recommendations Guidelines and request for Meta's review team to reassess the determination.

As to historical information, Meta does not maintain the information necessary to respond to this request. Additionally, to the extent this request calls for the disclosure of user content, we are prohibited from doing so by the Stored Communications Act.

Question 49. **What percentage of U.S.-based recommendations on your platform(s) are political in nature, such as accounts of political figures or content discussing current political issues? If you do not include political content in recommendations, please (a) elaborate on why not and (b) provide your precision rate for enforcing this rule.**

We want those who use our services to have a valuable experience across all of our platforms, which is why we personalize and recommend the content you see based on the choices you make. People have repeatedly told us they want to see less political content, so we have spent the last few years refining our approach on Facebook to reduce the amount of political content—including from politicians' accounts—people see in Feed, Reels, Watch, Groups You Should Join, and Pages You May Like. We have recently extended this approach in Reels, Explore, and In-Feed Recommendations on Instagram and Threads, too.

As part of this, and given the feedback we have received, we aim to avoid making recommendations that could be about politics or political issues, in line with our approach of not recommending certain types of content to those who do not wish to see it.

At the same time, we preserve the ability to find and interact with political content that is meaningful to you if that is what you are interested in on Facebook Feed. When ranking political content in Facebook Feed, our AI systems consider personalized signals, like survey responses, that help us understand what is informative, meaningful, or worth your time. We also consider how likely people are to provide us with negative feedback on posts about political issues when they appear in Facebook Feed. We have shifted away from ranking political content in Facebook Feed based on engagement signals—such as how likely you are to comment on or share content—since we have found that they are not reliable indicators that the content is of value or interest to someone.

Question 50. **Please list the top 100 sources of political content shown in recommendations, as defined by total distribution from recommendations, for each year over the past ten years. Please provide these lists regardless of whether you have a policy to not include political content in recommendations.**

Over the past few years, the amount of political content involved in recommendations—and therefore, sources of political content shown in recommendations—has been reduced.

People have repeatedly told us they want to see less political content, so we have spent the last few years refining our approach on Facebook to reduce the amount of political content – including from politicians' accounts—people see in Feed, Reels, Watch, Groups You Should Join, and Pages You May Like. We have recently extended this approach in Reels, Explore, and In-Feed Recommendations on Instagram and Threads, too. As part of this, and given the feedback we have received, we aim to avoid making recommendations that could be about politics or political issues, in line with our approach of not recommending certain types of content to those who do not wish to see it.

Question 51. **Please list all federal, state, and local elected officials that have been removed from or downranked in recommendations, even if temporarily, for a period of at least three continuous days within the past ten years. Please include the duration of and reason for the restriction, and note whether the restriction is currently in effect.**

Our Recommendations Guidelines provide information on when accounts, pages, or groups may be removed from recommendations on our platforms. These apply equally to the accounts of federal, state, and local elected officials. On Facebook, accounts, pages, or Groups that repeatedly violate our policies may be removed from recommendations and have their distribution reduced. On Instagram, recently or repeatedly posting policy-violating content can result in the entire account becoming ineligible to be recommended.

As a general matter, the amount of political content involved in recommendation has been reduced over the past few years.

People have repeatedly told us they want to see less political content, so we have spent the last few years refining our approach on Facebook to reduce the amount of political content – including from politicians' accounts—people see in Feed, Reels, Watch, Groups You Should Join, and Pages You May Like. We have recently extended this approach in Reels, Explore and In-Feed Recommendations on Instagram and Threads, too. As part of this, and given the feedback we have received, we aim to avoid making recommendations that could be about politics or political issues, in line with our approach of not recommending certain types of content to those who do not wish to see it.

Question 52. **What protocols do you have in place, if any, to audit the accuracy of your recommendation systems relative to your platform's stated rules?**

Our Recommendations Guidelines for Instagram and Facebook govern content that is eligible to be recommended. We have teams dedicated to maintaining, improving, and measuring the efficacy and reliability of our machine learning models that determine if content or accounts

meet our Recommendations Guidelines. We continue to invest in ensuring our recommendations systems are reliable. For instance, we have human reviewers review content against our Recommendations Guidelines to both improve the reliability of our machine learning models and validate their current performance.

Question 53. **How do you ensure that content, entities, and accounts are not being improperly or mistakenly filtered from your recommendation systems?**
Questions for the Record, Joint Hearing between Senate Commerce Committee and Senate Judiciary Committee on "Facebook, Social Media Privacy, and the Use and Abuse of Data" (April 10, 2018)

Mistakes are always possible, however, we have teams dedicated to maintaining, improving, and measuring the efficacy and reliability of our machine learning models that determine if content or accounts meet our Recommendations Guidelines. We continue to invest in improving the reliability of our recommendations systems. For instance, we have human reviewers review content against our Recommendations Guidelines to both improve the reliability of our machine learning models and validate their current performance.

We also provide feedback for account owners whose content is deemed ineligible to be recommended. If an account's content is not eligible to be recommended, the account owner can see a sample of content or components of the profile that may violate the Recommendations Guidelines and any content that has been removed for violating Meta's Community Standards. Users can also edit or delete posts that may violate the Recommendations Guidelines and request for Meta's review team to reassess the determination.

Question 54. **Yes or no: Does Facebook promote, demote, or block users or content based on its assessment of the social value or social desirability of that content?**

No. There are no policies specifically designed to do this, but Feeds are highly personalized to the interests of account holders.

The Facebook Community Standards outline what is and is not allowed on Facebook, and the Instagram Community Guidelines outline what is and is not allowed on Instagram. Meta takes a three-part approach to content enforcement on Facebook and Instagram: remove, reduce, and inform. We remove content that goes against our policies as soon as we become aware of it. Some problematic content can create a negative experience for people on Facebook and Instagram. We will often reduce the distribution of this content, even when it does not quite meet the standard for removal under our policies. When content is potentially sensitive or misleading, we sometimes add a warning prior to the user accessing the underlying content or share additional information from independent fact-checkers.

The goal of our policies is to create a place for expression and give people a voice. Meta wants people to be able to talk openly about the issues that matter to them, whether through written comments, photos, music, or other artistic mediums, even if some may disagree or find them objectionable. In some cases, we allow content—which would otherwise go against our standards—if it is newsworthy and in the public interest. We do this only after weighing the public interest value against the risk of harm, and we look to international human rights standards to make these judgments. In other cases, we may remove content that uses ambiguous or implicit language when additional context allows us to reasonably understand that the content goes against our standards.

Our commitment to expression is paramount, but we recognize the internet creates new and increased opportunities for abuse. For these reasons, when we limit expression, we do it in service of one or more of the following values: authenticity, safety, privacy, and dignity.

Specifically on the topic of content distribution, it is important to note that our ranking and recommendations systems are dynamic and highly complex and depend, in part, on the ever-changing interests across the billions of people who use our services. Across our services, we make personalized recommendations to the people who use our services to deliver new communities and content. Both Facebook and Instagram may recommend content, accounts, and entities that people do not already follow. Our goal is to make recommendations that are relevant and valuable to each person who sees them. For this reason, our recommendations are unique and highly personalized. Ultimately, we make recommendations based on content people have expressed interest in and actions they take on our apps.

Question 55. **Yes or no: Does Facebook promote, demote, or block users or content based on its assessment of that content's truth or falsity?**

We have a global network of fact-checking partners, all certified through the International Fact Checking Network, who independently review and rate potential misinformation across Facebook, Instagram, and WhatsApp. Their work enables us to take action and reduce the spread of problematic content across our apps.

Each time a fact-checker rates a piece of content as false on our platforms, we significantly reduce that content's distribution so that fewer people see it, label it accordingly, and notify people who try to share it. Fact-checkers do not remove content, accounts, or Pages from our apps. We remove content when it violates our Community Standards or Community Guidelines, which are separate from our fact-checking programs.

***Question 56*. Yes or no: Does Facebook promote, demote, or block users or content based on its assessment of the content's agreement or disagreement with Facebook's corporate values, beliefs, priorities, or opinions?**

No. The Facebook Community Standards outline what is and is not allowed on Facebook, and the Instagram Community Guidelines outline what is and is not allowed on Instagram. Meta takes a three-part approach to content enforcement on Facebook and Instagram: remove, reduce, and inform. We remove content that goes against our policies as soon as we become aware of it. Some problematic content can create a negative experience for people on Facebook and Instagram. We will often reduce the distribution of this content, even when it does not quite meet the standard for removal under our policies. When content is potentially sensitive or misleading, we sometimes add a warning prior to the user accessing the underlying content or share additional information from independent fact-checkers.

The goal of our policies is to create a place for expression and give people a voice. Meta wants people to be able to talk openly about the issues that matter to them, whether through written comments, photos, music, or other artistic mediums, even if some may disagree or find them objectionable. In some cases, we allow content—which would otherwise go against our standards—if it is newsworthy and in the public interest. We do this only after weighing the public interest value against the risk of harm, and we look to international human rights standards to make these judgments. In other cases, we may remove content that uses ambiguous or implicit language when additional context allows us to reasonably understand that the content goes against our standards. When we amend our content policies, we do so using a process that includes extensive engagement across a range of worldwide stakeholders and a review of external and internal research.

Our commitment to expression is paramount, but we recognize the internet creates new and increased opportunities for abuse. For these reasons, when we limit expression, we do it in service of one or more of the following values: authenticity, safety, privacy, and dignity.

Specifically on the topic of content distribution, it is important to note that our ranking and recommendations systems are dynamic and highly complex and depend, in part, on the ever-changing interests across the billions of people who use our services. Across our services, we make personalized recommendations to the people who use our services that deliver new communities and content. Both Facebook and Instagram may recommend content, accounts, and entities that people do not already follow. Our goal is to make recommendations that are relevant and valuable to each person who sees them. For this reason, our recommendations are unique and highly personalized. Ultimately, we make recommendations based on content people have expressed interest in and actions they take on our apps.

Question 57. **Yes or no: Have Facebook's decisions to permit users access to its services or to permit content to remain displayed on its services, or the prominence or accessibility of that content, including its order, visibility, duration visible, inclusion in searches or order within search results, inclusion within "Trending" lists or analogous suggestions of content to users, ever been determined in whole or part by Facebook's corporate values, beliefs, priorities, or opinions?**

Meta's policies, and its enforcement of those policies, govern user access to its services, what content appears on its services, and the order and visibility of content. We are clear and transparent about what our standards are, and we seek to apply them to all of our users consistently, regardless of any employee's personal values or preferences.

For example, our Recommendations Guidelines provide information on when accounts, pages, or Groups may be removed from recommendations on our platforms. In developing these guidelines, we sought input from 50 leading experts specializing in recommender systems, expression, safety, and digital rights. Those consultations are part of our ongoing efforts to improve these guidelines and provide people with a safe and positive experience when they receive recommendations on our platform. On Facebook, accounts, pages, or Groups that repeatedly violate our policies may be removed from recommendations and have their distribution reduced, affecting the prominence, visibility, and order of the content in the Feed. Enforcement of our Recommendations Guidelines involves having teams dedicated to maintaining and improving the efficacy and reliability of our machine learning models that determine if content or accounts meet our Recommendations Guidelines. We continue to invest in ensuring our recommendations systems are reliable. For instance, we have human reviewers review content against our Recommendations Guidelines to both improve the reliability of our machine learning models and validate their current performance.

Question 58. **Yes or no: Has Facebook ever discriminated among users on the basis of viewpoint when determining whether to permit a user to access its services? If so, please list each instance in which Facebook has done so.**

a. **If so, does Facebook continue to do so today, or when did Facebook stop doing so?**
b. **If so, what viewpoint(s) has Facebook discriminated against or in favor of? In what way(s) has Facebook done so?**
c. **If so, does Facebook act only on viewpoints expressed on Facebook, or does it discriminate among users based on viewpoints expressed elsewhere? Has Facebook ever based its decision to permit or deny a user access to its services on viewpoints expressed off Facebook?**

We moderate content according to our published policies, including our Facebook Community Standards and Instagram Community Guidelines. The political affiliation of the person generating the content has no bearing on content removal assessments.

***Question 59*. Yes or no: Excluding content encouraging physical self-harm, threats of physical violence, terrorism, and other content relating to the credible and imminent physical harm of specific individuals, has Facebook ever discriminated among content on the basis of viewpoint in its services? If so, please list each instance in which Facebook has done so.**

We moderate content according to our published policies, including our Facebook Community Standards and Instagram Community Guidelines. The political affiliation of the person generating the content has no bearing on content removal assessments.

***Question 60*. Yes or no: Has Facebook ever discriminated against American users or content on the basis of an affiliation with a religion or political party? If so, please list each instance in which Facebook has done so and describe the group or affiliation against which (or in favor of which) Facebook was discriminating.**

We moderate content according to our published policies, including our Facebook Community Standards and Instagram Community Guidelines. The political affiliation of the person generating the content has no bearing on content removal assessments.

***Question 61*. Yes or no: Has Facebook ever discriminated against any American users or content on its services on the basis of partisan affiliation with the Republican or Democratic parties? This question includes advocacy for or against a party or specific candidate or official. If so, please list each instance and the party affiliation discriminated against.**

We moderate content according to our published policies, including our Facebook Community Standards and Instagram Community Guidelines. The political affiliation of the person generating the content has no bearing on content removal assessments.

***Question 62*. Yes or no: Has Facebook ever discriminated against any American users or content on its services on the basis of the user's or content's advocacy for a political position on any issue in local, State, or national politics? This question includes but is not limited to advocacy for or against abortion, gun control, consumption of marijuana, and net neutrality.**

We moderate content according to our published policies, including our Facebook Community Standards and Instagram Community Guidelines. The political affiliation of the person generating the content has no bearing on content removal assessments.

***Question 63*. Yes or no: Has Facebook ever discriminated against any American users or content on its services on the basis of the user's or content's religion, including advocacy for one or more tenets of that religion? If so, please list each such instance in which Facebook has done so and identify the religion, religious group, or tenet against which Facebook discriminated.**

We moderate content according to our published policies, including our Facebook Community Standards and Instagram Community Guidelines. The political affiliation of the person generating the content has no bearing on content removal assessments.

***Question 64*. Yes or no: Has Facebook ever discriminated between users in how their content is published, viewed, received, displayed in "trending" or similar lists, or otherwise in any function or feature, based on the user's political affinity, religion, religious tenets, ideological positions, or any ideological or philosophical position asserted? If so, please list each such incident as well as the basis on which Facebook discriminated against that user or content.**

We moderate content according to our published policies, including our Facebook Community Standards and Instagram Community Guidelines. The political affiliation of the person generating the content has no bearing on content removal assessments.

In 2016, when allegations of political bias surfaced in relation to Facebook's Trending Topics feature, we immediately launched an investigation to determine if anyone violated the integrity of the feature or acted in ways that are inconsistent with Facebook's policies and mission. We spoke with current and former reviewers and their supervisors; reviewed our guidelines, training, and practices; examined the effectiveness of operational oversight designed to identify and correct mistakes and abuse; and analyzed data on reviewers' implementation of our guidelines.

Our investigation revealed no evidence of systematic political bias in the selection or prominence of stories included in the Trending Topics feature. We were unable to substantiate any of the specific allegations of politically motivated suppression of subjects or sources, as reported in the media. To the contrary, we confirmed that most of the subjects and sources identified were in fact included as trending topics on multiple occasions, on dates and at intervals that would be expected given the volume of discussion around the topics on relevant dates. For additional details on our investigation, please see our 2016 Newsroom Post, available at: https://about.fb.com/news/2016/05/response-to-chairman-john-thunes-letter-on-trending-topics/.

In 2018, we removed the Trending Topics feature from Facebook because we found that people no longer found it useful.

***Question 65*. Did or does Facebook collaborate with or defer to any outside individuals or organizations in determining whether to classify a particular statement as "hate speech?" If so, please list the individuals and organizations.**

While Meta consults with outside individuals and organizations on its approach to hate speech, including with outside academics and experts from across the political spectrum and around the world, Meta remains responsible for the development and enforcement of those policies. Meta's hate speech policies are laid out in our public Community Standards.

As a matter of policy, we do not share the names of the groups we consult with for a number of reasons, including safety and security concerns—which are especially acute in places around the world where the government may exercise censorship or control—and the fact that groups may not want to be named. The minutes from our Policy Forum, the process we use to develop our Hate Speech and other content policies, can be found here.

The Oversight Board has issued a number of opinions regarding Meta's Hate Speech policy, including just this past month when Meta reversed its original decision after the Board brought the appeal to our attention.

***Question 66*. Yes or no: Does Facebook contract with or in any way rely upon an outside party to determine what organizations and people are dedicated to promoting hatred against protected groups? If yes, please list the outside parties.**

While Meta consults with outside individuals and organizations on its approach to hate speech, including with outside academics and experts from across the political spectrum and around the world, Meta remains responsible for the development and enforcement of those policies. Meta's hate speech policies are laid out in our public Community Standards.

As a matter of policy, we do not share the names of the groups we consult with for a number of reasons, including safety and security concerns—concerns which are especially acute in places around the world where the government may exercise censorship or control—and the fact that groups may not want to be named. The minutes from our Policy Forum, the process we use to develop our Hate Speech and other content policies, can be found here.

***Question 67*. What percentage of Facebook's moderators:**

a. **Self-identify or are registered as Democrats?**

b. Self-identify or are registered as Republicans?
c. Would identify themselves as "liberal?"
d. Would identify themselves as "conservative?"
e. Have donated to:
1. **The Democratic Party?**
2. **A candidate running for office as a Democrat?**
3. **A cause primarily affiliated with or supported by the Democratic Party?**
4. **A cause primarily affiliated with or supported by liberal interest groups?**
5. **A political action committee primarily advocating for the Democratic Party, Democratic candidates or office-holders, or causes primarily supported by the Democratic Party?**
6. **The Republican Party?**
7. **A candidate running for office as a Republican?**
8. **A cause primarily affiliated with or supported by the Republican Party?**
9. **A cause primarily affiliated with or supported by conservative interest groups?**
10. **A political action committee primarily advocating for the Republican Party, Republican candidates or office-holders, or causes primarily supported by the Republican Party?**

f. Worked on or volunteered for a Democratic campaign?
g. Worked on or volunteered for a Republican campaign?
h. Worked [for], interned for, or volunteered for a Democratic legislator, State or federal?
i. Worked [for], interned for, or volunteered for a Republican legislator, State or federal?
j. Worked [for] or interned for a Democratic administration or candidate?
k. Worked [for] or interned for a Republican administration or candidate?

We do not maintain statistics on these data points.

Question 68. **What percentage of Facebook's employees:**
a. Self-identify or are registered as Democrats?
b. Self-identify or are registered as Republicans?
c. Would identify themselves as "liberal?"
d. Would identify themselves as "conservative?"
e. Have donated to:
1. **The Democratic Party?**
2. **A candidate running for office as a Democrat?**
3. **A cause primarily affiliated with or supported by the Democratic Party?**
4. **A cause primarily affiliated with or supported by liberal interest groups?**

5. **A political action committee primarily advocating for the Democratic Party, Democratic candidates or office-holders, or causes primarily supported by the Democratic Party?**
6. **The Republican Party?**
7. **A candidate running for office as a Republican?**
8. **A cause primarily affiliated with or supported by the Republican Party?**
9. **A cause primarily affiliated with or supported by conservative interest groups?**
10. **A political action committee primarily advocating for the Republican Party, Republican candidates or office-holders, or causes primarily supported by the Republican Party?**

f. **Worked on or volunteered for a Democratic campaign?**
g. **Worked on or volunteered for a Republican campaign?**
h. **Worked [for], interned for, or volunteered for a Democratic legislator, State or federal?**
i. **Worked [for], interned for, or volunteered for a Republican legislator, State or federal?**
j. **Worked [for] or interned for a Democratic administration or candidate?**
k. **Worked [for] or interned for a Republican administration or candidate?**

We do not maintain statistics on these data points.

Question 69. **What percentage of Facebook's management:**

a. **Self-identify or are registered as Democrats?**
b. **Self-identify or are registered as Republicans?**
c. **Would identify themselves as "liberal?"**
d. **Would identify themselves as "conservative?"**
e. **Have donated to:**
    1. **The Democratic Party?**
    2. **A candidate running for office as a Democrat?**
    3. **A cause primarily affiliated with or supported by the Democratic Party?**
    4. **A cause primarily affiliated with or supported by liberal interest groups?**
    5. **A political action committee primarily advocating for the Democratic Party, Democratic candidates or office-holders, or causes primarily supported by the Democratic Party?**
    6. **The Republican Party?**
    7. **A candidate running for office as a Republican?**
    8. **A cause primarily affiliated with or supported by the Republican Party?**
    9. **A cause primarily affiliated with or supported by conservative interest groups?**

10. **A political action committee primarily advocating for the Republican Party, Republican candidates or office-holders, or causes primarily supported by the Republican Party?**

f. **Worked on or volunteered for a Democratic campaign?**
g. **Worked on or volunteered for a Republican campaign?**
h. **Worked [for], interned for, or volunteered for a Democratic legislator, State or federal?**
i. **Worked [for], interned for, or volunteered for a Republican legislator, State or federal?**
j. **Worked [for] or interned for a Democratic administration or candidate?**
k. **Worked [for] or interned for a Republican administration or candidate?**

We do not maintain statistics on these data points.

Question 70. **What percentage of Facebook's executives:**

a. **Self-identify or are registered as Democrats?**
b. **Self-identify or are registered as Republicans?**
c. **Would identify themselves as "liberal?"**
d. **Would identify themselves as "conservative?"**
e. **Have donated to:**
    1. **The Democratic Party?**
    2. **A candidate running for office as a Democrat?**
    3. **A cause primarily affiliated with or supported by the Democratic Party?**
    4. **A cause primarily affiliated with or supported by liberal interest groups?**
    5. **A political action committee primarily advocating for the Democratic Party, Democratic candidates or office-holders, or causes primarily supported by the Democratic Party?**
    6. **The Republican Party?**
    7. **A candidate running for office as a Republican?**
    8. **A cause primarily affiliated with or supported by the Republican Party?**
    9. **A cause primarily affiliated with or supported by conservative interest groups?**
    10. **A political action committee primarily advocating for the Republican Party, Republican candidates or office-holders, or causes primarily supported by the Republican Party?**
f. **Worked on or volunteered for a Democratic campaign?**
g. **Worked on or volunteered for a Republican campaign?**
h. **Worked [for], interned for, or volunteered for a Democratic legislator, State or federal?**

i. **Worked [for], interned for, or volunteered for a Republican legislator, State or federal?**
j. **Worked [for] or interned for a Democratic administration or candidate?**
k. **Worked [for] or interned for a Republican administration or candidate?**

We do not maintain statistics on these data points.

***Question 71*. On what basis does Facebook evaluate whether to honor a foreign government's request to block specific content?**

We do not remove content simply because someone—even a government—requests it. When we receive a report or request to take down content, we evaluate the content against the Facebook Community Standards or Instagram Community Guidelines and, if we determine that the content goes against our policies, we take action. We may also restrict access to content that does not violate our policies but is alleged to violate local law, and if the report comes from a government entity, we first conduct a careful legal review and a human rights assessment.

A number of countries around the world have laws that limit content that might otherwise be allowed by our Community Standards or US law. In Germany, for example, laws forbid incitement to hatred. There are times when we may have to remove or restrict access to content because it violates a law in a particular country, even though it does not violate our Community Standards. Further, when governments believe that something on the internet violates their laws, they may contact companies like Meta and ask us to restrict access to that content. When we receive such a request, it is scrutinized to determine if the request has come from the recognized authority in the country and if the specified content does indeed violate local laws. If we determine that it does, then we may make it unavailable in the relevant country or territory.

In cases where we believe that reports are not legally valid, are overly broad, or are inconsistent with international human rights standards, we may request clarification or take no action. In all cases, we consider the impact our decisions will have on the availability of other speech via our services.

Where we do act against organic content on the basis of local law rather than our Community Standards, we restrict access to the content only in the jurisdiction where it is alleged to be unlawful and do not impose any other penalties or feature restrictions. We also notify the affected user. More information is available at https://transparency.fb.com/reports/content-restrictions/content-violating-local-law.

***Question 72*. How does Facebook determine whether to honor a foreign government's request to block specific content or users?**

We do not remove content simply because someone—even a government—requests it. When we receive a report or request to take down content, we evaluate the content against the Facebook Community Standards or Instagram Community Guidelines and, if we determine that the content goes against our policies, we take action. We may also restrict access to content that does not violate our policies but is alleged to violate local law, and if the report comes from a government entity, we first conduct a careful legal review and a human rights assessment.

A number of countries around the world have laws that limit content that might otherwise be allowed by our Community Standards or US law. In Germany, for example, laws forbid incitement to hatred. There are times when we may have to remove or restrict access to content because it violates a law in a particular country, even though it does not violate our Community Standards. Further, when governments believe that something on the internet violates their laws, they may contact companies like Meta and ask us to restrict access to that content. When we receive such a request, it is scrutinized to determine if the request has come from the recognized authority in the country and if the specified content does indeed violate local laws. If we determine that it does, then we may make it unavailable in the relevant country or territory.

In cases where we believe that reports are not legally valid, are overly broad, or are inconsistent with international human rights standards, we may request clarification or take no action. In all cases, we consider the impact our decisions will have on the availability of other speech via our services.

Where we do act against organic content on the basis of local law rather than our Community Standards, we restrict access to the content only in the jurisdiction where it is alleged to be unlawful and do not impose any other penalties or feature restrictions. We also notify the affected user. More information is available at https://transparency.fb.com/reports/content-restrictions/content-violating-local-law.

***Question 73*. Listed by country, what percentage of requests to block specific content (or users) from foreign governments does Facebook honor in whole or part?**

We report the number of pieces of content restricted in each country where our services are available. You may view information about how we restrict content based on local law here: https://transparency.fb.com/reports/content-restrictions/.

We do not remove content simply because someone—even a government—requests it. When we receive a report or request to take down content, we evaluate the content against the Facebook Community Standards or Instagram Community Guidelines and, if we determine that the content goes against our policies, we take action. We may also restrict access to content that does not

violate our policies but is alleged to violate local law, and if the report comes from a government entity, we first conduct a careful legal review and a human rights assessment.

A number of countries around the world have laws that limit content that might otherwise be allowed by our Community Standards or US law. In Germany, for example, laws forbid incitement to hatred. There are times when we may have to remove or restrict access to content because it violates a law in a particular country, even though it does not violate our Community Standards. Further, when governments believe that something on the internet violates their laws, they may contact companies like Meta and ask us to restrict access to that content. When we receive such a request, it is scrutinized to determine if the request has come from the recognized authority in the country and if the specified content does indeed violate local laws. If we determine that it does, then we may make it unavailable in the relevant country or territory.

In cases where we believe that reports are not legally valid, are overly broad, or are inconsistent with international human rights standards, we may request clarification or take no action. In all cases, we consider the impact our decisions will have on the availability of other speech via our services.

Where we do act against organic content on the basis of local law rather than our Community Standards, we restrict access to the content only in the jurisdiction where it is alleged to be unlawful and do not impose any other penalties or feature restrictions. We also notify the affected user. More information is available at https://transparency.fb.com/reports/content-restrictions/content-violating-local-law.

***Question 74*. How does Facebook determine whether to honor the U.S. government's request to block specific content or users?**

We do not remove content simply because someone—even a government—requests it. When we receive a report or request to take down content, we evaluate the content against the Facebook Community Standards or Instagram Community Guidelines and, if we determine that the content goes against our policies, we take action. We may also restrict access to content that does not violate our policies but is alleged to violate local law, and if the report comes from a government entity, we first conduct a careful legal review and a human rights assessment.

A number of countries around the world have laws that limit content that might otherwise be allowed by our Community Standards or US law. In Germany, for example, laws forbid incitement to hatred. There are times when we may have to remove or restrict access to content because it violates a law in a particular country, even though it does not violate our Community Standards. Further, when governments believe that something on the internet violates their laws, they may contact companies like Meta and ask us to restrict access to that content. When we

receive such a request, it is scrutinized to determine if the request has come from the recognized authority in the country and if the specified content does indeed violate local laws. If we determine that it does, then we may make it unavailable in the relevant country or territory.

In cases where we believe that reports are not legally valid, are overly broad, or are inconsistent with international human rights standards, we may request clarification or take no action. In all cases, we consider the impact our decisions will have on the availability of other speech via our services.

Where we do act against organic content on the basis of local law rather than our Community Standards, we restrict access to the content only in the jurisdiction where it is alleged to be unlawful and do not impose any other penalties or feature restrictions. We also notify the affected user. More information is available at https://transparency.fb.com/reports/content-restrictions/content-violating-local-law.

***Question 75*. What percentage of requests to block specific content (or users) from the U.S. government does Facebook honor in whole or part?**

Our Transparency Report contains data on restrictions we place on content that does not violate community standards but that is alleged to violate local law. We do not have any such reports for the United States. You may view the Transparency Report here: https://transparency.fb.com/reports/content-restrictions/country/US/.

Questions from Senator Grassley

Question 1. **Current law requires that a provider of a report of suspected CSAM to the National Center for Missing and Exploited Children's (NCMEC) CyberTipline preserve "any visual depictions, data, or other digital files that are reasonably accessible and may provide context or additional information about the reported material or person" for a minimum of 90 days. 18 U.S.C. 2258A(h)(1-2). The recent explosion of suspected abuse has presented unprecedented challenges for law enforcement to follow up on leads before companies discard or delete essential data and information. There is nothing preventing tech companies from preserving relevant material beyond the statutorily-mandated 90-day period.**

a. **How long does Meta voluntarily preserve and retain data contained in and related to its reports to the CyberTipline?**
b. **The massive influx of reports to the CyberTipline naturally results in law enforcement entities having to conduct and finish investigations beyond 90 days of an initial report to the CyberTipline. Retaining relevant information for longer periods could significantly advance law enforcement's ability to thoroughly investigate leads. If Meta only preserves and retains this information for the minimum 90-day period, why does it do so when preserving this data longer could significantly enhance and prolong law enforcement's ability to investigate and prosecute child predators?**
c. **Please confirm if Meta stores and retains the following information relating to reports to the CyberTipline:**
    i. **IP addresses**
    ii. **Screen Names**
    iii. **User Profiles**
    iv. **Associated Screennames (by IP address and associated emails)**
    v. **Email addresses**
    vi. **Geolocation data**
d. **If Meta does not retain or store any of the above types of information in question (c), please explain why.**
e. **Please list any other information Meta retains and preserves for law enforcement purposes not listed above in question (c).**
f. **Does Meta flag screennames and associated email addresses to suspected accounts that violate Meta's terms of service?**

We are proud of the strong relationship we have developed with NCMEC and continue to report all CSAM found globally to NCMEC's CyberTipline across our family of apps. We have built systems and review processes to prioritize and appropriately action violating content or accounts and, when appropriate, report it to NCMEC or law enforcement. NCMEC has

acknowledged Meta as an industry leader in this work and that Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs." We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

As a general matter, Electronic Service Providers are legally obligated to report apparent violations of laws related to child sexual abuse material they become aware of to NCMEC's CyberTipline. To do so, we submit electronic reports that contain the apparent child exploitative image(s). We endeavor to make our reports robust and include various types of information allowed by law in order to protect people and our services. Additionally, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations. We provide operational guidelines to law enforcement who seek records from Facebook or Instagram.

We support efforts to develop common industry standards on child exploitation, including standards related to Cybertips. In order to do that well it is important to understand that companies across the industry provide a wide variety of services, and as a result of these differences they have access to different types of information to include in these reports. Accordingly, each company's report may vary based on a variety of factors, including information available and accessible to each provider. Additionally, industry standards must balance the feasibility of more detailed robust reporting with the need for timely submissions.

In addition to reporting content we become aware of, we go beyond the legal requirements and use sophisticated technology to proactively seek out this content, and as a result we find and report more CSAM to NCMEC than any other service today. We make this technology available to the industry to help protect children from exploitation across the internet.

With respect to our cooperation with law enforcement, we have developed a streamlined online process through which we accept and review all legal requests from law enforcement. We expedite requests pertaining to child safety, and we have a team dedicated to engaging with NCMEC, Child Exploitation and Online Protection Command, Interpol, the FBI, and numerous other local, federal, and international law enforcement organizations and departments to help make sure that they have the information and training needed to make the best use of this process and that we are supporting efforts to improve these processes. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC to help safeguard the child.

We comply with federal law that requires reported content and certain additional information to be preserved for 90 days following the submission of a NCMEC report. We also comply with government requests to preserve account information pending our receipt of formal legal

process. We respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations. We will also provide metadata, including potentially critical location or account information. We provide operational guidelines to law enforcement who seek records from Facebook or Instagram.

We note that we disagree with the premise of this question that "there is nothing preventing tech companies from preserving relevant material beyond the statutorily-mandated 90-day period," if "relevant material" here refers to the retention of child sex abuse material, the possession of which is subject to criminal penalties. Without clear statutory parameters, industry risks being in violation of possession laws. We welcome progress in amending the current preservation statute (18 U.S.C. 2258A(h)) to enable a longer preservation period. A longer preservation period with protections for using that data to improve our detection efforts would also allow the industry to further develop technology to fight this type of content. The reality of the current system is that volumes of reports are high, matters are increasingly complex—often crossing platforms—and the limited statutory period may simply be insufficient to meet the purpose of protecting children.

We have strict policies against child nudity, abuse, and exploitation, including CSAM, inappropriate interactions with children, solicitation, and content that sexualizes children. We go beyond legal requirements and use sophisticated technology to find, remove, and report child sexual abuse material and disrupt the networks of criminals behind it. We disable accounts that appear to be involved in malicious distribution of CSAM or sexual solicitation of children, and report apparent instances of child exploitation identified on our site to NCMEC, which coordinates with law enforcement authorities from around the world. We report more CSAM to NCMEC than any other service today.

In addition to our policies, we use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts, and we do not recommend teen accounts to these accounts, or vice versa. We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations, and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

Question 2. **How does Meta prioritize urgent requests for information from law enforcement and what is Meta's response time to urgent requests?**

We work with law enforcement, and deeply respect and support the work agencies do to keep us safe. The amount of time it takes to respond to certain legal process may depend on a variety of factors. In all cases, we carefully review, validate, and respond to law enforcement requests, and we prioritize emergency situations, including terrorism and child abuse. We also reach out to law enforcement when we see a credible threat of imminent offline harm, contacting federal, state, or local law enforcement depending on the specific circumstances of a threat.

We have specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations review potentially violating content and report findings to the National Center for Missing and Exploited Children (NCMEC).The reports to NCMEC include content from around the world, and in turn, NCMEC works with US federal, state, and local law enforcement, as well as law enforcement globally, to find and help victims.

With respect to our cooperation with law enforcement, we have developed a streamlined online process through which we accept and review all legal requests from law enforcement. We expedite requests pertaining to child safety, along with other emergency situations, and we have a team dedicated to engaging with NCMEC, Child Exploitation and Online Protection Command, Interpol, the FBI, and numerous other local, federal, and international law enforcement organizations and departments to help make sure that they have the information and training needed to make the best use of this process and that we are supporting efforts to improve these processes. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC) to help safeguard the child.

We dedicate significant resources to addressing the concerns of law enforcement authorities and ensuring the timely processing of legal requests. We have law enforcement response teams available around the clock to respond to emergency requests.

Question 3. **What is Meta's average response time to service of legal process from law enforcement for CSAM-related information?**

We recognize that we have a responsibility to work with law enforcement, and we deeply respect and support the work law enforcement agencies do to keep us safe. The amount of time it takes to respond to certain legal process may depend on a variety of factors. In all cases, we carefully review, validate, and respond to law enforcement requests, and we prioritize emergency situations, including terrorism and child abuse. We also reach out to law enforcement when we see a credible threat of imminent offline harm, contacting federal, state, or local law enforcement depending on the specific circumstances of a threat.

We have specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations review potentially violating content and report findings to the National Center for Missing and Exploited Children (NCMEC).The reports to NCMEC include content from around the world, and in turn, NCMEC works with US federal, state, and local law enforcement, as well as law enforcement globally, to find and help victims.

With respect to our cooperation with law enforcement, we have developed a streamlined online process through which we accept and review all legal requests from law enforcement. We expedite requests pertaining to child safety, along with other emergency situations, and we have a team dedicated to engaging with NCMEC, Child Exploitation and Online Protection Command, Interpol, the FBI, and numerous other local, federal, and international law enforcement organizations and departments to help make sure that they have the information and training needed to make the best use of this process and that we are supporting efforts to improve these processes. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC to help safeguard the child.

We dedicate significant resources to addressing the concerns of law enforcement authorities and ensuring the timely processing of legal requests. We have law enforcement response teams available around the clock to respond to emergency requests.

Question 4. **In 2023, the tech industry as a whole slashed more than 260,000 jobs. And in the first four weeks of this year, another 25,000 jobs were cut.**

a. **For each year, between 2018 and 2023, how many U.S. based employees did you have at Meta?**

Meta publicly reports total headcount. For each year between 2018 and 2023, Meta reported the following number of employees, excluding contractors, in its global workforce:

a. 67,317, as of December 31, 2023

b. 86,482, as of December 31, 2022

c. 71,970, as of December 31, 2021

d. 58,604, as of December 31, 2020

e. 44,942, as of December 31, 2019

f. 35,587, as of December 31, 2018

i. Of these employees, how many were sponsored on H-1B visas?

ii. For each year, between 2018 and 2023, how many H1-B visa applications did Meta submit?

Meta endeavors to comply with all applicable immigration laws in the United States and the other countries where we operate. The Department of Labor publicly reports employment-based immigration data on the agency's website: https://www.dol.gov/agencies/eta/foreign-labor/performance.

b. For each year, between 2018 and 2023, how many employees based outside the U.S. did you have at Meta?

i. Of these employees, how many were based in China?

Meta maintains significant international operations. We currently make Facebook available in more than 100 different languages, and we have offices or data centers in approximately 40 different countries.

c. For each year, between 2018 and 2023, how many employees in total did Meta terminate, fire, or lay off?

i. Of these employees, how many were based in the United States?

ii. Did Meta fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?

iii. Were any duties and/or functions previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?

After years of growth, Meta implemented a company-wide restructuring plan focused on flattening our organization. The goal of these efforts was to make the company faster, leaner, and more efficient. To be clear, these restructuring efforts did not change the commitment we have to our ongoing integrity efforts. We have brought teams together to think across a number of key issues. For example, our Global Operations team now works more closely with our integrity team, and we have consolidated certain support teams from different areas across the company.

To be clear, we absolutely remain committed to our work keeping people safe on our services. Even with the targeted changes, we continue to have about 40,000 people focused on overall safety and security efforts. Finding efficiencies in our work has been a focus for years. We will continue to hire across security and integrity teams to support our industry-leading work in the most efficient and effective manner possible.

d. For each year, between 2018 and 2023, how many employees performing work related to child safety did Meta terminate, fire, or lay off?
   i. Of these employees, how many were based in the United States?
   ii. Did Meta fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?
   iii. Were any duties and/or functions (specifically relating to child safety) previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?
   iv. How have layoffs impacted Meta's ability to protect children on its platforms?
   v. Does Meta have any plans to increase staff responsible for child safety operations or otherwise optimize its child safety operations?

Please see the response to your Question 4(c).

***Question 5*. On January 30, 2024, the Tech Transparency Project (TTP) published an article on their website called, "Meta Approves Harmful Teen Ads with Images from its Own AI Tool". In summary, TTP, using Meta's "Imagine with Meta AI" tool generated inappropriate images such as young people at a pill party or other vaping. These images with text were submitted to Facebook as advertisements targeting users between ages 13-17 in the United States. TTP reported that Facebook approved the advertisement, despite it violating its own policies, in less than five minutes to run on the following platforms: Facebook, Instagram, Messenger, and Meta Quest. Meta. Over the course of a week, TTP submitted the advertisements with the same end result: Facebook approving them. TTP reported that they canceled these advertisements before their scheduled publication, but it illustrated the repeated failures of Facebook to properly moderate content. This is just one example of what other non-government organizations and others have uncovered across social media platforms.**

a. How often a month do Meta employees conduct quality checks on Meta's policies and safeguards for child accounts?
b. In which departments, components, or units of the company does Meta have staff dedicated to performing this type of work?
c. How many employees make up these departments, components, or units?
d. If a violation is found, what action is taken, and how quickly is action taken?

Keeping young people safe online has been a challenge since the start of the internet. That is why now, as much as ever, we are working hard to stay ahead by working with specialists dedicated to online child safety and sharing information with our industry peers. People want to use a service that makes them feel safe. That means we are incentivized to prioritize safety and it is why we have invested in building integrity solutions, and continually review our policies and procedures to address safety and security on our services.

Since 2016, Meta has significantly expanded the number of people who work on overall safety and security. By 2018, Meta doubled the number of people who work on safety issues from 10,000 to 20,000, which includes content reviewers, systems engineers, and security experts. By 2020, Meta built a global team of 35,000 people to work on safety and security. And by 2022, Meta had more than quadrupled the number of people working on safety and security since 2016 to over 40,000 people. We continue to have around 40,000 people devoted to safety and security efforts.

Regarding the TTP reporting referenced, as we have seen with other generative AI models across the industry, it is possible for AI features to share inaccurate or inappropriate outputs. We block terms like "child," "teen," or "youth" when combined with certain requests, and are continually working to improve our blocklists. For example, we have blocklists to prevent our tools from returning images based on certain additional prompts like "young + vape" or "young + rifle."

We also stress test our services to improve safety performance and collaborate with policymakers, experts in academia and civil society, and others in our industry to advance the responsible use of this technology. We take a number of steps to identify potential vulnerabilities, reduce risks, and enhance the safety of our generative AI features, including:

- Evaluating and improving AIs with external and internal experts;
- Fine-tuning our models and training them to provide expert-backed resources in certain contexts (e.g., in response to queries about suicide); and
- Developing new technology to catch and take action on violating policies.

No AI model is perfect. We will use the feedback we receive to keep training the models to improve safety performance and automatic detection of policy violations.

Question 6. **Social media companies claim they are investing in company components dedicated to safety, and that their platforms are safe for children. However, children continue to be exploited daily across these platforms.**

a. **What have Meta's revenue and profit figures been for the last three years (2021-2023)? Please provide figures broken out per year. Do not provide percentages.**
b. **How much has Meta spent in advertising for the last three years (2021-2023), broken out per year?**
c. **How much of Meta's resources spent on advertising has been devoted to advertising Meta's safety initiatives and efforts for the last three years (2021-2023), broken out per year?**

Meta discloses revenue and net income figures in our public SEC filings, which are reproduced below.

Revenue:

- For the twelve months ended on December 31, 2023, Meta reported $134.90 billion in revenue.
- For 2022, Meta reported $116.61 billion in revenue.
- For 2021, Meta reported $117.93 billion in revenue.

Net Income:

- For the twelve months ended on December 31, 2023, Meta reported a net income of 39.1 billion.
- For the twelve months ended on December 31, 2022, Meta reported a net income of 23.2 billion.
- For the twelve months ended on December 31, 2021, Meta reported a net income of 39.4 billion.

As a general matter, we do not share detailed descriptions of our spending allocations for specific advertising campaigns. We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts. That is why we have invested more than $20 billion overall in safety and security since 2016, and we will never stop working on these issues.

d. **To get an understanding of how your company has invested and plans to invest in its components dedicated to child safety functions, what are the annual budgets for Meta's child safety-related components for the last three years (2021-2023)?**

We are unable to provide a more precise estimate of different child safety related budgets, as this work is embedded throughout the company. For a more detailed breakdown of Meta's child safety-related components, which operate across our various teams and partnerships, please see the response to your Question 6(i).

We have around 40,000 people overall working on safety and security, including on child safety functions, and we have invested over $20 billion since 2016. This includes around $5 billion in 2023. In 2022, Meta invested approximately $6 billion on safety and security. In 2021, Meta invested about $5 billion on safety and security.

e. What is the current anticipated (2024) budget for Meta's child safety-related components?

Please see the response to your Question 6(d).

f. Provide the number of staff employed in Meta's child safety-related components for the last three years (2021-2023).

Please see the response to your Question 6(d).

Since 2016, Meta has significantly expanded the number of people who work on overall safety and security. By 2018, Meta doubled the number of people who work on safety issues from 10,000 to 20,000, which includes content reviewers, systems engineers and security experts. By 2020, Meta built a global team of 35,000 people to work on safety and security. And by 2022, Meta had more than quadrupled the number of people working on safety and security since 2016 to over 40,000 people.

g. How much is that compared to Meta's other components for the same period? (Please provide a breakout per year. Do not provide percentages.)

Please see the response to your Question 6(f).

h. How many staff are currently employed in Meta's child safety-related components?

Please see the response to your Question 6(d).

i. What are the roles, responsibilities, and functions of Meta's child safety-related components?

Child safety is extremely important to us. We have specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to review potentially

violating content and report findings to the NCMEC. We work to find, remove, and report child sexual abuse material and disrupt the networks of criminals behind it. We developed technology that identifies adult accounts exhibiting potentially suspicious behavior, reviewing a number of signals to proactively find and restrict potential predators. We deploy machine learning to proactively detect accounts engaged in certain suspicious patterns of behavior by analyzing dozens of combinations of metadata and public signals, such as if a teen blocks or reports an adult. When we identify these accounts, we work to limit their ability to find, follow, or interact with teens or each other, and we automatically remove them if they exhibit a number of these signals.

As required by law, we report all apparent instances of child exploitation identified on our site from anywhere in the world to NCMEC, which coordinates with law enforcement authorities from around the world. We respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations. Between 2020 and 2023, our teams disrupted 37 abusive networks and removed nearly 200,000 accounts associated with those networks. In Q4 2023, we removed 16.2 million pieces of child sexual exploitation content on Facebook and 2.1 million pieces on Instagram. In Q4 2023, of the child sexual exploitation content we actioned, we detected 99% on Facebook and 95% on Instagram before it was reported by our users.

Child safety components also work with industry partners to better keep kids safe. Child protection requires a global and comprehensive response from industry, law enforcement, government, civil society, and families, which is why we are committed to working with child-safety stakeholders to build and support the child-safety ecosystem. We also collaborate across the industry through organizations like the Technology Coalition, an industry association dedicated solely to eradicating the sexual exploitation of children online. In 2020, Meta joined Google, Microsoft, and 15 other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual abuse. This project includes a renewed commitment and investment from the Technology Coalition, expanding its scope and impact to protect kids online and guide its work for years to come. Project Protect focuses on five key areas: tech innovation, collective action, independent research, information and knowledge sharing, and transparency and accountability. We also collaborate with industry and law enforcement on new programs, such as Lantern. Lantern is a program that enables technology companies to share signals of a child safety threat and Take It Down is a platform designed to proactively prevent young people's intimate images from spreading online.

Additionally, we work closely with safety advisors, which include leading online safety nonprofits, as well as over 400 safety experts and NGOs from around the world, including specialists in combating child-sexual exploitation and aiding its victims. Our efforts include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, among other things.

j. Are any other components responsible for the monitoring of CSAM on Meta's platform(s)?

Please see the response to Question 6(i) for a description of Meta's CSAM monitoring processes, technologies, and partnerships.

k. What, if any, third parties does Meta employ or contract with to address CSAM material on its platforms?
   i. What are the roles and responsibilities of these third parties?
   ii. What is the breakdown of cost per third party over the last three years (2021-2023)?

We work regularly with child safety professionals to help us understand evolutions in coded language and to identify new and evolving terms, phrases, slang, and emojis that could be used in an attempt to evade our detection systems and bypass our policies.

We also work with these professionals and organizations to build various interventions, including but not limited to our search interventions, safety notices, and safety education campaigns. We have also worked with child safety researchers to conduct collaborative research to improve child safety protections on our platforms.

Our collaborative work to address child safety does not stop with improving our own services. We also are deeply committed to improving the entire ecosystem and have engaged with child safety nonprofits and academic researchers to complete child safety research with fieldwide impact. Our efforts with these professionals also include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, among other things. Additionally, to help safety stakeholders identify and respond to the most high priority reports at NCMEC, we funded and helped rebuild their case management tool to ensure investigators can get to the most important cases quickly.

Because this is an industry-wide concern, we also direct people to various tools to use if people have nude or sexual photos or videos to help prevent them from being shared or reshared online. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

On Take It Down, young people or their guardians can submit a case to proactively search for attempted uploads of their intimate images on participating platforms. Take It Down allows people to only submit a hash—rather than the intimate image or video itself—to NCMEC. Hashing turns images or videos into a coded form that can no longer be viewed, producing hashes that are secure digital fingerprints. Once a person submits the hash to NCMEC, companies like ours can use those hashes to find any copies of the image, evaluate any matches of images attempting to be uploaded to confirm they violate our policies, block the upload, and help prevent the content from being posted on our platforms in the future—helping to return power and control back to the victim.

Anyone seeking support and information related to sextortion can visit our education and awareness resources, including the Stop Sextortion resources we developed with Thorn. These resources include immediate actions parents and teens can take if users are experiencing sextortion, as well as expert tips for teens, parents and guardians, and information for parents on how to talk to their teens about intimate images. We also worked with Thorn and their NoFiltr brand to create and promote educational materials that reduce the shame and stigma surrounding the sharing of intimate images, and empower teens to seek help and take back control if they have shared them or are experiencing sextortion.

Additionally, we collaborate across the industry through organizations like the Technology Coalition, an industry association dedicated solely to eradicating the sexual exploitation of children online. In 2020, Meta joined Google, Microsoft, and 15 other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual abuse. This project includes a renewed commitment and investment from the Technology Coalition, expanding its scope and impact to protect kids online and help guide its future work. Project Protect focuses on five key areas: tech innovation, collective action, independent research, information and knowledge sharing, and transparency and accountability. We also announced our recent participation in Lantern, a Tech Coalition program that enables technology companies to share a variety of signals about accounts and behaviors that violate their child safety policies. Lantern participants can use this information to conduct investigations on their own platforms and take action. Meta was a founding member of Lantern, providing the Tech Coalition with the technical infrastructure that sits behind the program and encouraging our industry partners to use it. We manage and oversee the technology with the Tech Coalition, ensuring it is simple to use and provides our partners with the information they need to track down potential predators on their own platforms.

Question 7. **Of all reports sent by Meta to the National Center for Missing and Exploited Children, how many reports were self-generated from victim users for the last three years (2021-2023)? Please provide the actual number of self-generated reports in addition to the**

total number of reports (including those that were not self-generated). In addition, please provide a break-down of the self-reporters by age.

We encourage everyone to report anything they think may violate our policies. We have made our reporting tools easier to find and started encouraging teens to report at relevant moments, such as when they block someone. We do not keep the statistics requested, as reports to NCMEC can take many forms on our services, such as through proactive detection and reactive reporting by users.

We have invested heavily in different tools to help increase user reports. For example, on Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Because this is an industry-wide concern, we also direct people to various tools to use if people have nude or sexual photos or videos to help prevent them from being shared or reshared online. Instagram and Facebook are founding members of Take It Down—a platform by NCMEC to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Question 8. **The proliferation of end-to-end encryption is expected to result in a sharp reduction in reports of suspected child abuse to NCMEC. A natural foreseeable result of this could be significant cost reductions to Meta, enabling the company to allocate staff resources elsewhere as there will be less reviewable material for Meta staff to analyze before sending reports to NCMEC. Given that Meta is responsible for the vast majority of reports to NCMEC, Meta's transition to default end-to-end encryption is expected to result in a massive decrease in reports to the CyberTipline and the vast majority of CSAM on Meta's platforms going undetected and thus unreported to law enforcement (given children's understandable hesitation to report their own abuse).**

a. **Does Meta anticipate that its transition to default end-to-end encrypted messaging across its platforms will result in lower costs related to child safety operations and resources devoted to detecting and reporting CSAM on its platforms? If so, what is the anticipated amount of cost-savings?**

No. We do not anticipate a lowering in spend related to child safety operations and resources in 2024 as a result of our transition to end-to-end encryption or otherwise. Our rollout of end-to-end encryption is not driven by cost savings, but by our belief that end-to-end encryption is one of the strongest tools we have to protect the privacy and security of people and their messages, which include some of our most personal communications shared with friends and family. As our lives move more and more online, we believe it is critical to preserve a space for private conversations where people can have the freedom to be themselves and share their most personal thoughts with loved ones. We have built systems and review processes to prioritize and appropriately action violating content or accounts and, when appropriate, report it to NCMEC or law enforcement. As discussed further in response to your Question 8(b), we expect to continue to focus on this work, and we expect this will continue to result in us providing more reports to law enforcement than our peers.

b. As Meta further implements end-to-end encryption on its platforms, is Meta developing or exploring technology that could potentially detect and report CSAM material in end-to-end encrypted messages sent on its platform?

As we expand encryption to Messenger and Instagram Direct Messages, our approach to safety is focused on three key elements: (i) preventing potential harm in the first place; (ii) giving people ways to control their experience; and (iii) responding to violations of our policies quickly. This approach is detailed in our whitepaper, Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging. In line with this, we are continually developing and exploring technology to assist in child safety in the context of encrypted services.

In an end-to-end encrypted environment, we use machine learning to proactively detect accounts engaged in potentially malicious patterns of behavior. Our machine learning technology will look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect potentially suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors. To help us respond to violations of our policies quickly, we also encourage people to report messages to us in both encrypted and unencrypted services.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself,

nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

The technology we explore and develop is designed in accordance with our Security Design Principles. For this reason we have not adopted, and do not intend to develop, scanning technologies that automatically access and report messaging content in end-to-end encrypted messages, often called "client-side scanning." These types of technologies, whether on a person's device or otherwise, without that person's consent and control could be abused by online criminals, malicious hackers or authoritarian regimes, putting people's safety at risk. We do not believe such technologies can be developed and implemented in a manner that is rights-respecting, nor can such technologies meet the expectations people have of end-to-end encrypted messaging services, and significant security concerns have been raised by leading technical experts in the field.

We also work with professionals, collaborate with industry, and support law enforcement around the world to fight the online exploitation of children. For example, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations. We expect to continue providing more reports to law enforcement than our peers, thanks to our industry-leading work on keeping people safe. For example, WhatsApp—which has long been encrypted—takes action against hundreds of thousands of accounts every month for suspected child exploitative imagery sharing. In 2022, WhatsApp also made over one million reports to NCMEC, all without breaking encryption. This was significantly more than all other encrypted

messaging services combined. NCMEC has acknowledged Meta continues to be an industry leader in this work and that Meta "goes above and beyond to make sure that there are no portions of their network where this type of activity occurs."

End-to-end encryption is already widely used by other large messaging services to protect people's private messages and provide people with the privacy and security they expect when messaging friends and family. We do not believe moving to an encrypted messaging environment means sacrificing safety. That is why we will continue to support encryption, while putting features in place to help keep people safe. We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

c. Do Meta's platforms currently provide the option for users to opt out of default end-to-end encryption? If not, why?

For private messaging, we will not offer the ability to opt out of default end-to-end encryption. People should have secure, private places where they have clear control over who can communicate with them and confidence that no one else can access what they share. Communications that are end-to-end encrypted reinforce safety and security and have become the standard user expectation for their preferred communications platforms. Creating an option to opt out of encryption would weaken the overall security of our private messaging services because messages would no longer be guaranteed to be private, and could confuse users who expect a safe and secure messaging experience.

End-to-end encryption by default is an important baseline functionality for private messages because it enables people to communicate in ways that are private and secure, while also simplifying the experience of people who use our platform. We follow a set of core security principles to ensure that security is central to the design of our messaging apps. With opt-out encryption, there is a risk that some messages may not be encrypted, leaving them vulnerable to interception. Vulnerable people who use our services do not always know that they are vulnerable, due to changing regulations or personal situations, or may just forget to encrypt sensitive chats. For more information, and to review our core security principles for messaging, visit https://engineering.fb.com/2022/07/28/security/five-security-principles-for-billions-of-messages-across-metas-apps/.

d. Does Meta offer or intend to offer a concerned parent the ability to decide whether or not their child's messages should be end-to-end encrypted? If not, why?

We want teens to have safe, age-appropriate experiences on our apps, including on our encrypted services. To help protect teens from unwanted contact, we have built tools and policies

specifically to help young people manage interactions with adults. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists. We also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

In addition to the investments described in response to your previous questions, we have also built more than 50 tools, resources, and features to help support teens. Additionally, parental supervision tools on Messenger allow parents to see how their teen uses Messenger, including how much time they are spending on messaging and information about their teen's message settings. Parents can also: view and receive updates on their teen's Messenger contacts list, as well as their teen's privacy and safety settings; get notified if their teen reports someone (if the teen chooses to share that information); view who can message their teen (only their friends, friends of friends, or no one) and see if their teen changes this setting; and can view who can see their teen's Messenger stories and get notified if these settings change.

Meta also offers Messenger Kids, a parent-controlled messaging service for children under 13 years old. The service is not encrypted, because the ability for parents to view messages is not compatible with end-to-end encryption.

***Question 9*. What is Meta's policy or protocol with respect to law enforcement accessing user data and subsequent notification to users of law enforcement accessing their data?**

a. **Do certain crimes such as drug trafficking or child exploitation affect Meta's decision to notify a user whose data is accessed by law enforcement?**
b. **Do certain requests such as a subpoena or search warrant affect Meta's notification protocol? If so, what are they?**

c. If Meta does notify users of law enforcement accessing their data, why does Meta find this necessary?

We disclose account records solely in accordance with our terms of service and applicable law, including the federal Stored Communications Act ("SCA"), 18 U.S.C. Sections 2701-2712. Accordingly, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations. We provide operational guidelines to law enforcement who seek records from Facebook or Instagram.

Our policy is to notify people who use Facebook and Instagram of requests for their information prior to disclosure unless we are prohibited by law from doing so or in exceptional circumstances, such as child exploitation cases, emergencies or when notice would be counterproductive. Law enforcement officials who believe that notification would jeopardize an investigation can, and often do, obtain an appropriate court order or other appropriate process establishing that notice is prohibited.

Question 10. **The National Center for Missing and Exploited Children has indicated that reports from social media companies tend to lack actionable information causing law enforcement to be burdened with incomplete information. How comprehensive are Meta's reports to NCMEC? What challenges is Meta experiencing on the collection of user data and other information to include in its reports to NCMEC? What actions is Meta taking to make its reports more comprehensive and useful to law enforcement?**

We are proud of the strong relationship we have developed with NCMEC and continue to report all apparent CSAM found globally to NCMEC's CyberTipline across our family of apps. We have built systems and review processes to prioritize and appropriately action violating content or accounts and, when appropriate, report it to NCMEC or law enforcement. NCMEC has acknowledged that Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs."

As a general matter, Electronic Service Providers are legally obligated to report apparent violations of laws related to child sexual abuse material they become aware of to NCMEC's CyberTipline. To do so, we submit electronic reports that contain the apparent child exploitative image(s). We endeavor to make our reports robust and include various types of information allowed by law in order to protect people and our services.

In addition to reporting content we become aware of, we go beyond the legal requirements and use sophisticated technology to proactively seek out this content, and as a result we find and report more CSAM to NCMEC than any other service today. We make this technology available

to the industry to help protect children from exploitation across the internet. For example, we find and report far more content to NCMEC than any other internet service today. In 2022, all of the industry made 32 million reports to NCMEC collectively. We made over 26 million reports between Facebook and Instagram. The rest of the industry made less than 6 million reports collectively.

Frequently, our systems *immediately* detect CSAM being uploaded, enabling us to thwart the attempt to distribute the content. Even in such instances, we submit a report, which reflects not incidences of child sexual exploitation on our platforms, but rather *failed* attempts by bad actors to use our platform for this abhorrent crime. Our automation has enabled us to report at high volumes because we detect at scale, which in turn drives the volume of reporting. We will continue to refine our systems, and we call upon the rest of the industry to do the same.

We will continue collaborating with organizations like NCMEC and child safety experts, while working to prevent the spread of CSAM online.

***Question 11*. The Metaverse has been described as the next iteration of the internet. Technology companies have invested significantly into the new platform. On January 1, 2024, the Daily Mail published an article that British police are investigating a virtual gang rape of a girl under the age 16 in the Metaverse. The girl, during a virtual realty video game, was attacked by several adult men. While the underage girl did not experience a physical attack, law enforcement reported that the girl suffered the same psychological and emotional trauma as someone who has been raped in the real world. Also, the article reports that there have been a number of reported sex attacks in Horizon Worlds, one of Meta's free VR (virtual reality) online games.**

a. **Based on this article, what proactive steps is Meta taking to ensure people, especially children, do not experience rape, assault, or any unwanted sexual advances?**
b. **How can you ensure no other person experiences this psychological and emotional trauma?**

Teens have become fans of popular virtual experiences across the industry—this makes it crucial that we build age-appropriate and positive experiences for them in virtual reality (VR). Doing so is core to our responsible innovation principles and our efforts to build safer experiences for young people.

Meta Horizon Worlds is one app among many multiplayer experiences in VR. Like other experiences available on our platform, Meta Horizon Worlds is a social experience in which people can explore, play, and create worlds with others. We have welcomed teens into Meta Horizon Worlds, where they can spend meaningful time with friends who may be separated by distance, explore their interests and passions, and express themselves. We have additional

protections and tools in place to help provide age-appropriate experiences, including strong defaults for features like voice mode, Meta Horizon profile privacy settings, and active status settings, discussed in more detail below.

We know it is important for people to feel safe and in control of their experience and surroundings in Meta Horizon Worlds. To that end, we have built features into the app that help empower people, including:

- Pause, a space where you can take a moment away from other people and your surroundings.
- The ability to block and mute people.
- The ability to report people or content in real time.
- A personal boundary so other players cannot get too close, which makes it easier to avoid unwanted interactions.

We have supported teens with safety tools and built-in protections to facilitate an age-appropriate experience since opening up Meta Horizon Worlds to them, including:

- **Limiting interactions between teens + adults they do not know:** We want to help prevent teens from hearing from adults they may not know. That is why we take steps to help prevent interactions between adults and unconnected teens. For example, we do not display any adults a teen does not know in their "people you might know" list.
- **Meta Horizon profile privacy settings:** Teens are in control of who they follow and who can follow them back. Teens' profiles are automatically set to private, so they are able to approve or decline anyone who requests to follow them.
- **Active status settings:** By default, we do not show a teen's active status and Meta Horizon Worlds location to other people in Worlds. Teens are able to choose whether their connections can see if they are active and which public world or event they are in.
- **World and event content ratings:** We use content ratings to help ensure teens have age-appropriate experiences within Worlds. For example, mature world and event ratings prevent teens from finding, seeing, or entering spaces that contain mature content. Our policies prohibit teens from publishing mature worlds or events. Worlds violating this policy are removed.
- **Voice mode:** This feature transforms the voices of people a teen does not know into quiet, friendly sounds, giving teens more control over who can communicate with them. It will also garble the teen's voice, so people they do not know can not hear them. We turn garbled voices on automatically for all teens by default within voice mode.

- **Blurred chats:** This feature is turned on by default for teens and obscures messages from people the teen doesn't know in world chat, and the teen's messages appear blurred to them, too. Parents can also view, adjust, and lock this feature within Parental Supervision.
- **World chat filter:** This filter automatically hides words or phrases that might be upsetting or offensive in world chat. World chat filter is turned on by default for all people, including teens. Parents can also view, adjust, and lock this feature within Parental Supervision.

We also prohibit certain content that people using the app might find offensive. In addition to what is not allowed under the Code of Conduct for Virtual Experiences, we do not allow the following in Meta Horizon Worlds:

- Content that depicts or promotes the use of illegal drugs or abuse of prescription drugs.
- Content that promotes criminal or dangerous activity.
- Content that attempts to buy, sell or trade real life regulated goods, such as firearms, blades, alcohol and tobacco.

Additionally, everyone—including teens—can cast their experience from Meta Quest, allowing them to share what they're seeing in VR and Worlds with parents, guardians, or others around them.

Further, Meta works with over 500 women's safety NGOs around the world through regional roundtables and during the UN Commission on the Status of Women to get ongoing feedback on our safety tools and collaborate with civil society experts to make all online spaces welcoming and safe for everyone.

Questions from Senator Hawley

Question 1. **Do you allow your children to use social media? If so, please explain under what conditions.**

Mark Zuckerberg's children are all under nine years old and use age-appropriate social media under his supervision. His eldest daughter uses Messenger Kids, a parent-managed service designed for children, on which parents and guardians can manage and control their child's experience. Meta believes it is important that if anyone is building a service for kids under the age of 13 to use, that there are appropriate parental controls.

Question 2. **Do you believe that children under the age of 18 should be allowed to use social media?**

Yes. Teens do amazing things on our services, and we are committed to helping teens have safe and positive experiences on them. Technology companies should build experiences that meet young people's needs while also working to keep them safe, and we are deeply committed to doing industry-leading work in safety and security. That is why we have invested more than $20 billion in safety and security across our platforms since 2016.

We have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

Question 3. **How many individuals does your company employ in Trust & Safety?**

Child safety efforts remain a top priority, which is why we continue to have around 40,000 people devoted to safety and security efforts.

Question 4. **How many individuals does your company employ to review content for so-called "misinformation," "disinformation," or "malinformation"?**

We generally use "misinformation" to mean the sharing of false content without intent to manipulate, and we generally use the term "disinformation" for the sharing of information with the intent to manipulate. Our work in these areas is reflected in our content moderation efforts, misinformation policies, third-party fact-checking program, and efforts to fight coordinated inauthentic behavior. Our integrity efforts remain a top priority, which is why we continue to have around 40,000 people devoted to safety and security efforts. This includes over 15,000 reviewers across the globe who review potential violations of our policies on Facebook and Instagram.

Since 2016, we have built an advanced system combining people and technology to review the billions of pieces of content that are posted to our platform every day. Our AI systems flag content that may violate our policies, users report content to us they believe is questionable, and our own teams review content. We also partner with over 90 fact-checking organizations around the world who rate content in more than 60 languages. At any given moment, many people are involved in identifying, labeling, or removing misinformation from our platforms and they all depend on one another to do it at the scale at which we operate.

When it comes to disinformation, we tackle it through our policies and enforcements against coordinated inauthentic behavior (CIB), which covers coordinated networks that centrally rely on fake accounts to mislead people about who they are and what they are doing to manipulate or corrupt public debate for a strategic goal. We conduct our own independent investigations and enforce against CIB. We do so based on the deceptive behavior we see on our platform, not based on the content they share. Our team focused on disrupting influence operations includes experts across the company, with backgrounds in law enforcement, national security, investigative journalism, cybersecurity, law, internet freedom, human rights, and engineering. Our technical teams continue to build scaled solutions to help detect and prevent these violating behaviors, and we work with civil society organizations, researchers, and governments to strengthen our defenses. We have also improved our detection systems to more effectively identify and block fake accounts, which are the source of a lot of inauthentic activity.

Question 5. **How many dollars per year does your company spend on salaries for Trust & Safety officers?**

While Meta does not publicly report salary information except where required by applicable law, we spent around $5 billion on safety and security in 2023.

Question 6. **Do you believe that the algorithms your company has developed to sort users' feeds are protected by Section 230 of the Communications Decency Act of 1995? If so, please explain why.**

Yes. At Meta, we believe in giving people a way to express themselves, while working hard to keep people safe across our services. The people and advertisers who use our services expect us to do this so we can continue to provide the most useful and engaging experience for them. One way Meta helps people to build community is by building and training algorithms to recommend connections and content people might be interested in—for example, new Facebook Groups they might want to join, Pages they might like, or events they might want to attend—and by ranking content so that they are more likely to see the posts they care most about. This technology also helps protect our community by filtering, blocking, and reducing the spread of content that violates our policies or is otherwise problematic.

The sheer volume of user-generated content on the internet means that online services have to make decisions about how to organize, prioritize, and deprioritize this content in ways that are useful to people and advertisers, while enforcing our policies against harmful content. Meta has invested billions of dollars to develop sophisticated safety and security systems that work to identify, block, and remove harmful content quickly—typically before it is ever seen by anyone. Section 230 was enacted to allow companies to do exactly this.

At a high level, Section 230 does two things. First, it encourages free expression by barring claims against online services for publishing third party speech. Without Section 230, online services could potentially be held liable for everything people say. Without this protection, such services may be likely to remove more content to avoid legal risk and may be less likely to invest in technologies that enable people to express themselves in new ways. Second, it allows online services to remove certain objectionable content. Without Section 230, such services could face liability, for example, for removing bullying and harassment content.

Question 7. **Do you believe that the algorithms your company has developed to sort users' feeds are expressive speech protected by the First Amendment to the U.S. Constitution? If so, please explain why.**

As discussed in the response to your Question 6, we take steps to organize, display, and disseminate third-party content to those who are most likely to be interested in it. We also use algorithms to remove objectionable content from our services. These activities are protected under the First Amendment.

Question 8. **Is your company a member of a party, an amicus, or a member of an amicus in NetChoice, LLC v. Paxton, No. 22-555 (U.S.), or did your company provide any funds or donations to any party or amicus in that case? If so, please describe the amount of funds or donations made and the context.**

Along with almost 40 other companies, Meta is a member of NetChoice and the Computer & Communications Industry Association, the petitioners in *NetChoice, LLC v. Paxton*. In addition, we supported the litigation in *NetChoice, LLC v. Paxton*. Meta belongs to various trade groups and organizations representing diverse views and communities. We also work with independent third-party organizations on issues relating to technology and internet policy, and we sometimes support their events that highlight internet and social media issues. We seek to participate in conversations about the issues that directly affect our company and the experience of the people who use our service. We chose these organizations because they are engaged in meaningful dialogue about either the internet or the local communities in which we operate. While we actively participate in these discussions and believe collaborative problem solving is the best way to address a problem and have the greatest impact, we do not always agree with every policy or position that individual organizations or their leadership take. Therefore, our membership, work with organizations, or event support should not be viewed as an endorsement of any particular organization or policy.

Question 9. **Do you believe that the First Amendment to the U.S. Constitution precludes Congress from enacting legislation holding social media companies liable to users for torts they commit?**

As a company, we have faced, currently face, and will continue to face claims and government inquiries relating to information or content that is published or made available on our services, including claims and inquiries relating to our policies, algorithms, and enforcement actions with respect to such information or content. We will continue to assert valid legal defenses to such claims as appropriate. The extent of those defenses will vary depending on the particularities of the claims. As we regularly do, we would welcome the opportunity to work with Congress on thoughtful legislative proposals.

Question 10. **Do you believe that companies can be trusted to develop artificial general intelligence (AI) through open-source methods?**

Yes. Exploratory research, open science, and cross-collaboration are foundational to Meta's AI efforts, and we have experienced first-hand how innovation in the open leads to technologies that advances the industry. For example, PyTorch has become one of the leading platforms for AI research as well as commercial production use with over 18,000 organizations using PyTorch. At Meta, PyTorch powers 50 on-device AI models across different mobile applications.

At Meta, we also believe that open sourcing models helps create safer services. By democratizing access to AI technology, potential vulnerabilities can be continuously identified and mitigated in a transparent way by an open community. We believe that openness will lead to better services, faster innovation, and a flourishing market, which benefits us as it does many

others. Open LLMs make it possible for businesses to participate and advance the AI industry without large amounts of funds, computing resources, or technical expertise. While it is true that businesses can use "closed" models (like OpenAI's GPT-4), building AI applications on top of closed models means that the developer will forever be beholden to the model developer, who will continue to receive usage fees and continue to exercise control over the infrastructure upon which the applications rely. Open models, by contrast, provide businesses with the foundation they need to innovate more quickly, free of charge (businesses must of course, comply with a set of usage guidelines and license terms). This means more independent innovators, less gate-keeping, more competition, and ultimately a more diverse AI industry.

In addition, open sourcing is a longstanding, well-regarded approach to enhancing security. In our experience, instead of creating more new risks than benefits, open source releases have helped us, and the broader community of developers, build safer and more robust systems. By democratizing access, vulnerabilities are continuously identified and mitigated by an open community, and that creates safer products.

As the FTC itself explained, "open-source pre-trained models" like our Llama models can help prevent "a market where the highest quality pre-trained models are controlled by a small number of incumbents" particularly when the open-source models available are of relatively high quality.[28] Open models can therefore "open up the playing field" towards quicker advancement and innovation. We also understand that there are risks with opening access to this foundational technology—and we have taken several steps so people can use our Llama models responsibly, including by restricting access to Meta's webform in certain jurisdictions.

We also understand the importance of building safeguards into AI tools from the beginning so that people can have safer, and ultimately more enjoyable, experiences. As AI technology continues to evolve, safety features and controls will also have to evolve. That is why we work in collaboration with stakeholders across industry and academia to make sure that AI systems have responsible guardrails. By open sourcing our AI tools, we aim to work and collaborate across industry, academia, government, and civil society to help ensure that such technologies are developed responsibly and with openness to minimize the potential risks and maximize the potential benefits. Any final decision on when, whether, and how to open source is taken following safety evaluations we run prior to launch.

Question 11. **Do you believe the government should play a role in licensing certain artificial intelligence technologies, such as generative AI products?**

As discussed in response to your Question 10, exploratory research, open science, and cross-collaboration are foundational to Meta's AI efforts. To promote long term advancement,

[28] https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns

we believe that openness will lead to better services, faster innovation, and a flourishing market, which benefits us as it does many others. For more on our commitment to open sourcing, please see the response to your Question 10.

Question 12. **Do you believe that artificial intelligence represents an existential threat to humanity?**

No. Though artificial intelligence technologies have great promise and potential, it is important to keep in mind that they do have limitations. While they can unlock a host of new possibilities in industries, from health care to logistics to manufacturing, they also have limited abilities to reason, a prevalent feature of human intelligence. As the technology exists today, even the most powerful AI systems are quite far from approximating human intelligence. Like all foundational technologies, there will be a multitude of uses of AI technologies, some predictable and some less so, which can be alarming. And like every technology, AI will be used by people for good and bad ends.

New technology brings new challenges, and everyone has a part to play here. Companies should make sure tools are built and deployed responsibly. We have invested in the responsible development of AI technology for more than a decade because we believe that AI has the potential to bring immense benefits to humanity. This investment has enabled us to play a significant role in identifying and addressing existing and emerging societal challenges. Meta's AI tools have aided in new scientific discoveries to improve environmental resilience, identified and increased coverage for social protection programs, and improved the way the world communicates through mass scale and accurate translation tools, among many other applications.

As we deepen our investment in AI technology, we will continue to consider how to develop and deploy these technologies responsibly. While AI has brought—and will continue to bring—huge advancements to society, we recognize that it comes with risks and the potential to cause unintended consequences. To ensure AI tools are built and used in a way that promotes the most beneficial outcomes for our society, we need consensus on the responsible ways to develop AI technology openly. Government, industry, academia, and civil society must work together to produce common and harmonized AI governance models, including globally agreed upon codes of practice, standards, and guardrails. That is why we are working to help advance the responsible design and operations of AI technology and are committed to building this technology thoughtfully from the start.

While more universal standards for advanced AI are still being established, we are developing our AI products and services with commitments to safety, security, and trust. Our responsible AI efforts are propelled by a cross-disciplinary team whose mission is to help ensure that AI at Meta benefits people and society. These efforts include work with datasets, balancing privacy and

fairness, helping to prevent bias in ad delivery systems, mitigating harmful or disrespectful associations, giving people more control over what they see, offering more transparency into AI models, and collaborating on standards and governance.

Question 13. **Do you think that the development of large language models by Microsoft, Google, Meta, and other large companies raises antitrust concerns?**

In February 2023, Meta announced the availability of Llama 2, the next generation of its open source LLM. We believe an open approach to AI allows a wide range of stakeholders—developers, academics, civil society, nonprofits, and more—to both realize the benefits of AI technologies and improve our understanding of how to manage and mitigate the potential risks. The addition of open models, like Llama 2, not only represents additional choice, but it also facilitates entry for smaller innovators and startups, allowing those who may not have the capital or resources to compete in this space. Giving businesses, startups, entrepreneurs, and researchers access to tools developed at a scale that would be challenging to build themselves, backed by computing power they might not otherwise access, will open up a world of opportunities for them to experiment, innovate in exciting ways, and ultimately benefit from economically and socially. For this reason, we believe projects like Llama 2 will only bring more competition to AI.

Question 14. **What steps does your company take to make transparent the algorithms by which users are censored, shadow banned, or demonetized?**

Meta has taken concrete steps to enhance understanding regarding its algorithms. An approach to transparency Meta has been developing and advocating for some time is the publication of system cards, which give people insight into how our systems work in a way that is accessible for those who do not have deep technical knowledge. We have released a number of system cards for Facebook and Instagram to date. They give information about how our AI systems rank content, some of the predictions each system makes to determine what content might be most relevant to you, as well as the controls you can use to help customize your experience. They cover Feed, Stories, Reels, and other surfaces where people go to find content from the accounts or people they follow. The system cards also cover AI systems that recommend "unconnected" content from people, groups, or accounts they do not follow.

We have also shared the types of inputs—known as signals—as well as descriptions of the predictive models these signals inform that help determine what content you will find most relevant from your network on Facebook. The categories of signals we have released represent the vast majority of signals currently used in Facebook Feed ranking for this content. You can find these signals and predictions in the Transparency Center, along with how frequently they tend to be used in the overall ranking process.

We also have made it possible to see details directly in our apps about why our systems predicted content would be relevant to you, and the types of activity and inputs that may have led to that prediction. We have expanded our "Why Am I Seeing This?" feature in Instagram Reels tab and Explore, and Facebook Reels, after previously launching it for some Feed content and all ads on both Facebook and Instagram. People are able to click on an individual Reel to see more information about how their previous activity may have informed the machine learning models that shape and deliver the reels they see.

Question 15. **What steps does your company take to ensure that your company is not disproportionately targeting or censoring conservative voices?**

Freedom of expression is a founding principle for Meta, and giving people a voice to express themselves has been at the heart of everything we do. We are committed to designing our apps to foster the free flow of ideas and culture, regardless of political affiliation. In a mature democracy with a free press, political speech is a crucial part of how democracy functions. And it is arguably the most scrutinized form of speech that exists.

We recognize both the importance of speech and how many can use speech to try to silence opposition and bully those who disagree with them. Our goal has been and remains to give everyone a voice, while also taking necessary steps so that our platforms remain safe spaces. Our Community Standards on Facebook and our Community Guidelines on Instagram include restrictions around content that is harmful to members of our community, including bullying, harassment, hate speech, and incitement to violence.

We moderate content according to our published policies in an effort to help keep people on our platforms safe, reduce objectionable content, and ensure people participate on our platforms responsibly. We strive to be clear and transparent about what our standards are, and we seek to apply them to everyone consistently. The political affiliation of the person generating the content has no bearing on content removal assessments. Decisions about whether to remove content are based on whether the content violates our Community Standards. Preventing people from seeing what matters most to them is directly contrary to our mission and our business objectives.

Question 16. **Do you condemn Hamas' terrorist attacks on the State of Israel on October 7, 2023?**

Yes. The terrorist attacks by Hamas were pure evil. There is never any justification for carrying out acts of terrorism, and we condemn them in the strongest possible terms. Meta has long considered Hamas to be a terrorist organization and the group is banned from our platforms.

People who use Facebook and Instagram are also prohibited from glorifying, supporting, or representing Hamas.

Question 17. **What role do you believe social media companies have in promoting or limiting public speech regarding the events of October 7, 2023?**

Since the onset of the current conflict, there has been a surge in related content on our platforms. We recognize Meta's role in responding to this intense crisis while seeking to keep human rights principles, and respect for civilians, at our core. While our platforms are designed to give everyone a voice, we also work to protect the safety and well-being of our community—and to respond to adversarial coordinated behaviors. In addition, Meta is the only tech company to have publicly released human rights due diligence on Israel-Palestine related issues. Our response to this conflict builds on that project, which we published in 2022 and updated in 2023.

Expert teams from across our company have been monitoring our platforms, while protecting people's ability to use our services to shed light on important developments happening on the ground. This includes enforcing our policies, which we apply regardless of who is posting or their personal beliefs. The balance between voice and safety is often not easy to strike in peaceful contexts. In conflict situations—and especially conflict situations involving sanctioned entities, such as Hamas—it is much more difficult. We know that people who use our apps, particularly Arabic and Hebrew speakers, have felt deeply impacted by our decisions. Some people think we take down too much content, while others think we remove too little. Ultimately, we seek to enforce our policies consistently and in alignment with our standards.
In some cases, we allow otherwise policy-violating content when its public interest value outweighs the risk of harm. We conduct a thorough assessment of any potentially newsworthy content and our reviewers consider a number of factors prior to escalating to our Content Policy team. We assess whether that content surfaces an imminent threat to public health or safety, or gives voice to perspectives currently being debated as part of a political process. We also consider other factors, such as country-specific circumstances, the nature of the speech, and the political structure of the country. To date, we have granted very limited exceptions for content related to the Israel-Hamas War.

We take our content moderation policies and enforcement very seriously, investing heavily to try and get it right. We recognize both the importance of speech and how many can use speech to try to silence opposition and bully those who disagree with them. While we know not everyone will agree with every decision and policy we make, we remain committed to providing transparency to our content moderation and enforcement policies.

Question 18. **What investments has your company made in anti-CSAM technology?**

Over the years, we have invested heavily in sophisticated technology that helps us proactively find violating content and accounts of this kind and remove them. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in enqueuing certain content for human review.

For example, we proactively detect and take action against known child exploitation and sexualizing content, leveraging technology available across industry for CSAM hash matching, including methods that Meta developed and open sourced. We also detect novel, previously unknown child exploitation and sexualizing content using proprietary machine learning detection technology, in conjunction with a team of specialized human reviewers. More specifically, we use PhotoDNA and other photo- and video-matching technologies that detect identical or near-identical photos and videos of known child exploitative content, and we use Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess. We also use technology to detect and remove Instagram Reels and Stories that violate our Community Guidelines, including by scanning for CSAM and for CSE indicators.

Additionally, we use a combination of technology and behavioral signals to detect and prevent potentially inappropriate interactions between minors and adults. We also use technology to bolster the systems we use to prioritize reports for content reviewers. For example, we are using technology designed to proactively find child exploitative imagery to identify and prioritize reports of content that are more likely to contain content that violates our child safety policies.

We also rely on both automated technology, reports, and investigations to take action on violating hashtags, account names, search terms, and emojis. We work to avoid showing search results for inherently violating terms (as well as terms that are not inherently violating, but that may be used by adversarial actors seeking or offering inappropriate content) to help prevent the discovery of potentially harmful content. Because we recognize this is a constantly evolving area, we also work with our specialist child safety teams and child safety professionals to help us understand evolutions in coded language and to identify new and evolving terms, phrases, slang, and emojis that could be used in an attempt to evade our detection systems and bypass our policies. Our teams use these signals and technology to proactively find new trends, misspellings and spelling variations of this language, as well as terms and phrases related to child exploitation, that we can input into our systems to proactively find and disrupt efforts to evade our protections.

We also use technology to find relationships between terms that we already know could be harmful or that break our rules and other terms used at the same time. These could be terms

searched for in the same session as violating terms, or other hashtags used in a caption that contains a violating hashtag. We combined our systems so that as new terms are added to our central list, they will be actioned across Facebook and Instagram simultaneously. We may send Instagram accounts, Facebook Groups, Pages, and Profiles to content reviewers, restrict these terms from producing results in Facebook and Instagram Search, and block hashtags that include these terms on Facebook and Instagram.

In addition to investments in our own technology, since 2019, we have also made two technologies—PDQ and TMK-PDQF—publicly available which detect identical and nearly identical photos and videos. We use PhotoDNA and other photo- and video-matching technologies that detect identical or near-identical photos and videos of known child exploitative content, and we use Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess. We also announced our recent participation in Lantern, a Tech Coalition program that enables technology companies to share a variety of signals about accounts and behaviors that violate their child safety policies. Lantern participants can use this information to conduct investigations on their own platforms and take action. Meta was a founding member of Lantern, providing the Tech Coalition with the technical infrastructure that sits behind the program and encouraging our industry partners to use it. We manage and oversee the technology with the Tech Coalition, ensuring it is simple to use and provides our partners with the information they need to track down potential predators on their own platforms. Finally, Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Question 19. **Have you read the Fifth Circuit's opinion in Missouri v. Biden, No. 23-30445?**

Meta's legal experts review relevant case law as necessary.

Question 20. **Do you dispute any factual findings in the Fifth Circuit's opinions or the district court's opinions?**

We moderate content on our apps according to our policies to help keep people on our platforms safe, reduce objectionable content, and enable people to participate on the platform responsibly. We seek to be clear and transparent about what our standards are and to apply them consistently.

We also seek to be transparent about the various ways in which we engage with government agencies and law enforcement. And we regularly release transparency reports that provide

information on our responses to government requests for data, content restrictions based on local law, and our enforcement of our Community Standards.

***Question 21*. Does your platform continue to receive requests from federal agencies to censor or promote certain content?**

We do not remove content simply because someone—even a government—requests it. When we receive a report or request to take down content, we evaluate the content against the Facebook Community Standards or Instagram Community Guidelines and, if we determine that the content goes against our policies, we take action. We may also restrict access to content that does not violate our policies but is alleged to violate local law, and if the report comes from a government entity, we first conduct a careful legal review and a human rights assessment.

In cases where we believe that reports are not legally valid, are overly broad, or are inconsistent with international human rights standards, we may request clarification or take no action. In all cases, we consider the impact our decisions will have on the availability of other speech via our services.

Where we do act against organic content on the basis of local law rather than our Community Standards, we restrict access to the content only in the jurisdiction where it is alleged to be unlawful and do not impose any other penalties or feature restrictions. We also notify the affected user. More information is available at https://transparency.fb.com/reports/content-restrictions/content-violating-local-law.

***Question 22*. What steps do your platforms take to verify and enforce age restrictions?**

Meta recognizes the need to keep people who are too young off Facebook and Instagram. Both Facebook's Terms of Service and Instagram's Terms of Use in the United States require people to be at least 13 years old to sign up for Facebook and Instagram. Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

Identifying Potentially Underage Accounts

Anyone, including individuals who do not have a Facebook or Instagram account, can report to Meta that someone is or appears to be under the age of 13 by filling out a platform-specific online reporting form. Reported accounts are evaluated either by an automated process or through human review, and in some cases both. Our content reviewers are trained to confirm and remove accounts that appear to be used by people who are underage. While manual review is a labor- and time-intensive process, Meta has taken steps to review accounts flagged as potentially underage as quickly as possible after they are reported. Additionally, Meta may learn that someone is underage directly from the individual, if the person attempts to change the date of birth on their account to a date that would make them under 13. In this case, the individual is automatically placed in an "age checkpoint," and we remove the account if the person cannot verify they are over 13.

Automated Evaluation

An account that has been flagged as potentially underage will first go through an automated process that determines whether the account should be escalated for human review or immediately allowed to continue using the platforms. Where Meta has evidence indicating that the reported individual is over the age of 13, Meta may automatically permit the person who has been flagged as potentially underage to continue using Facebook or Instagram. For example, this can occur when a human reviewer previously evaluated the account for potential underage usage and approved the individual to continue using the platform following the review (pursuant to review guidelines detailed below), the account was previously placed in an age checkpoint and the person submitted sufficient documentation demonstrating they were at least 13 years old, or the account is so old that it could not reasonably belong to a person under 13. A flagged account will also be permitted to remain on the platform if the account contains no bio or photos, because, as discussed below, a reviewer relies on this data to evaluate whether the account belongs to an underage person.

Manual Review of Potentially Underage Accounts

Flagged accounts that cannot be resolved through the automated processes described above are directed to human reviewers for further evaluation. Meta employs tens of thousands of human reviewers whose duties include reviewing these Facebook and Instagram accounts to manually look for signs that an account has violated the applicable terms of service or content guidelines, including accounts suspected as belonging to people under 13.

All potentially underage accounts that are manually reviewed are evaluated to determine whether they meet our age requirements. For example, reviewers consider the following factors:

- **Account Bio**: Reviewers first evaluate the account's bio for contextual information or self-admission about a person's real age, including a written statement of the person's age, birth date, or grade in school. Reviewers are also trained to look for additional signals that indicate whether the account holder is underage. An account that contains information that explicitly states or contextually implies that the individual is under 13 will be checkpointed and the person will be required to provide Meta with proof of age.
- **Account Photos**: If the account bio does not contain sufficient written information to determine age, the reviewer will evaluate the photos contained in the account. If a human reviewer finds sufficient signals that the account holder may be under 13, or is unsure of whether an account holder is under 13 based on a review of the account media, the account will immediately be placed in an age checkpoint.

Responding to Potentially Underage Accounts

When Meta identifies a potentially underage account, their account will be placed in an age checkpoint. While in the checkpoint, a person does not have access to their account, and they are shown a blocking screen if they attempt to log into their account. This means checkpointed accounts cannot view or interact with any content or ads on the platform. Additionally, checkpointed accounts are not visible to other people on the platform, and people cannot see or interact with the checkpointed account or the photos or videos posted to it.

If the person is unable to demonstrate that they are 13 years of age or older, their account is permanently disabled and removed, and the data is deleted consistent with Meta's standard deletion policies.

Other Mechanisms for Identifying Potentially Underage Accounts

In addition, we have also partnered with Yoti, a company that offers privacy-preserving ways to verify age. Yoti is verified by the Age Check Certification Scheme and is the leading age verification provider for several industries around the world including social media, gaming and age-restricted e-commerce. Expert and governmental organizations in youth and privacy have publicly endorsed Yoti for their approach and expertise in responsible artificial intelligence.

For example, on both Instagram and Facebook, a person who attempts to change their date of birth to go from the age of under 18 to 18 or older is required to verify their age through one of two options, ID upload or video selfie provided by the third-party Yoti. If Yoti estimates that the person is under the age of 13, the account will be placed in an age checkpoint. As explained above, if the person is unable to demonstrate that they are 13 years of age or older, their account is permanently disabled and removed, and when the account is disabled, the data is deleted consistent with Meta's standard deletion policies.

Question 23. **In cases where a child's safety is at risk, how does your company collaborate with law enforcement? What information or assistance is provided?**

With respect to our cooperation with law enforcement, we have developed a streamlined online process through which we accept and review all legal requests from law enforcement. We expedite requests pertaining to child safety, along with other emergency situations. We have a team dedicated to engaging with NCMEC, Child Exploitation and Online Protection Command, Interpol, the FBI, Homeland Security Investigations and numerous other local, federal, and international law enforcement organizations and departments to help them have the information and training needed to make the best use of this process and that we are supporting efforts to improve these processes. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC to help safeguard the child.

Question 24. **Do you believe there is any expressive value in CGI or AI generated CSAM?**

No. Child exploitation is a horrific crime that we work aggressively to fight on and off our services. We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. As with all of our policies, our prohibition against CSAM also applies to AI-generated material. We have processes in place to remove policy-violating content, regardless of the context or the person's motivation for sharing it. We also have developed aggressive, cutting-edge technology to help prevent, find, and remove policy violating content. Additionally, when we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children (NCMEC), in compliance with applicable law. In addition to this technology, we have invested in specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to both find and review potentially violating content, accounts and adversarial networks.

Question 25. **Do you believe that CGI or AI generated CSAM is protected by the First Amendment to the U.S. Constitution?**

No. We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material, inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children, in compliance with applicable law. For more information about our work fighting this content on and off our services, please see the response to your Question 24.

Question 26. **What measures does your platform take to ensure that children only see age-appropriate advertisements?**

We want everyone who uses our services to have safe, positive, and age-appropriate experiences. We have restricted the options advertisers have to reach teens, as well as the information we use to show ads to teens. Currently, age and geography are the only information about a teen that we use to show them ads. Last year, we made changes to how advertisers can reach teens, which included removing the ability for advertisers to target teens based on their gender, interest, and activities. We use age to show ads to teens because it helps us continue to show ads meant for their age. We use geography to comply with varying state laws. Additionally, we prohibit people under the age of 13 on any services that run advertising. And we do not show any ads in Messenger Kids, our messaging app for people under 13.

We also limit the kinds of ads that may be sent to teens. Our Advertising Standards prohibit ads about restricted topics—like alcohol, financial products and weight loss products and services—to be shown to people under 18 (and older in certain countries). Ads targeted to minors must not promote products, services, or content that are inappropriate, illegal, or unsafe, or that exploit, mislead, or exert undue pressure on the age groups targeted. Any ads promoting sexual and reproductive health products or services, like contraception and family planning, must be targeted to people 18 years or older and must not focus on sexual pleasure. This includes hotlines that appear on the landing pages of ads. If we detect that someone is under a certain age and attempts to view a Page or account with an age restriction, they will be blocked from viewing it. We have removed many ads for violating our policies around offering adult products and services.

In addition to the restrictions we have in place, we also provide teen-specific ad controls. Teens are able to manage the types of ads they see on Facebook and Instagram with Ad Topic Controls. As noted above, our Advertising Standards already prohibit ads about restricted topics—like alcohol, financial products and weight loss products and services—to be shown to people under 18 (and older in certain countries). But even when an ad complies with our policies, teens may want to see fewer ads like it. For example, if a teen wants to see fewer ads about a genre of TV show or an upcoming sports season, they should be able to tell us that. Teens can continue to choose to hide any or all ads from a specific advertiser. The topics we already restrict in our policies will be defaulted to See Less, so that teens cannot choose to opt into content that may not be age-appropriate. In addition to these controls, we have also introduced more teen-specific resources to help teens understand how ads work and the reasons why they see certain ads on our apps. These changes reflect research and direct feedback from parents and child developmental experts.

Question 27. **Will you commit to setting up a compensation fund for those who have been harmed by your platform?**

We want everyone who uses our services to have safe, positive, and age-appropriate experiences, and we approach all our work on child safety and teen mental health with this in mind. We build comprehensive controls into our services, we work with parents, experts, and teens to get their input, and we engage with Congress about what else needs to be done. We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone. We have built and shared tools for removing content that violates our policies, and we look at a wide range of signals to detect policy-violating behavior.

We are committed to protecting young people from abuse on our services, but this is an ongoing challenge. As we improve defenses in one area, criminals shift their tactics, and we have to come up with new responses. We will continue working with parents, experts, industry peers, and Congress to try to improve child safety, not just on our services, but across the internet as a whole.

Question 28. **You have stated that you want to build artificial general intelligence and do so through open source. What safeguards are you putting in place to ensure that the Chinese Communist Party does not repurpose this technology for nefarious purposes?**

We believe that it is important for the United States, and, specifically, US companies, to lead the way in creating the foundational AI tools and models that will be used the world over. It is key that the United States and other countries that share our values set the standard. At Meta, we also believe that open sourcing models creates safer services. In addition, open sourcing is a longstanding, well-regarded approach to enhancing security. In our experience, instead of creating more new risks than benefits, open source releases have helped us, and the broader community of developers, build safer and more robust systems. By democratizing access, vulnerabilities are continuously identified and mitigated by an open community, and that creates safer products.

Any final decision on when, whether, and how to open source is taken following safety evaluations we run prior to launch. In keeping with our commitment to Responsible AI development, Meta has undertaken a number of initiatives intended to discourage improper uses of its models. For example, we designed a bespoke license that includes a detailed and thought-out set of use restrictions that strictly prohibit a wide range of malicious uses.

Our Acceptable Use Policy clearly states that users may not "[c]reate, generate, or facilitate the creation of malicious code, malware, computer viruses or do anything else that could disable, overburden, interfere with or impair the proper working, integrity, operation or appearance of a

website or computer system." Additionally, we have implemented numerous ways of reporting violations of this policy including through reporting issues with the model, reporting risky content generated by the model, reporting bugs and security concerns or reporting violations of the Acceptable Use Policy.

We can use this information to take enforcement actions against individual licensees who violate our Acceptable Use Policy or who fail to comply with audits. Such enforcement actions may result in suspension or termination of the licensee's access to Llama models and/or referring violations to law enforcement. Meta is still working through an enforcement program, and we can keep you updated as the program evolves.

Question 29. **What safeguards are you putting in place to ensure that bad actors do not leverage open source artificial intelligence to produce or distribute child pornography?**

Please see the responses to your Questions 24 and 28.

Question 30. **What safeguards are you putting in place to ensure that bad actors do not leverage open source artificial intelligence to produce or distribute deep fake pornography?**

We publish Community Standards and Community Guidelines governing the types of content and behaviors that are acceptable on Facebook and Instagram. These policies apply to all content on these platforms, including content generated by AI. We prohibit pornography and sexually explicit content on our services, as well as offers or asks for pornographic material (including, but not limited to, sharing of links to external pornographic websites). We work to remove images that depict incidents of sexual violence and intimate images shared without the consent of the person(s) pictured. We also prohibit derogatory sexualized manipulated imagery of real people. When we find this content, we work to remove it, regardless of how it is created.

Meta has dedicated significant resources to detecting content on our services, including AI-generated content, that violates our policies. Our investments have allowed us to build technologies to help proactively identify potentially violating content, prioritize critical content for review, and act on content that violates our policies. We enforce our policies through a combination of people and technology that work to identify violations of our Community Standards across the billions of pieces of content that are posted to our services every day. For example, our systems flag content that may violate our policies, people who use our apps report content to us they believe is questionable, and our own teams review content. We work to remove content that violates our policies quickly and at scale, with the help of media-matching technology to find content that is identical or near-identical to photos, videos, text, and even audio that we have already removed. We have also built a parallel content review system to flag posts that may be going viral—no matter what type of content it is—as an additional safety net. This helps us catch content that our traditional systems may not pick up. We use this tool to help

detect and review Facebook and Instagram posts that were likely to go viral and take action if that content violated our policies.

We also strive to use a number of protections in our generative AI, including:

- **Training Our Model to Recognize Exploitative Queries**: We are training our models to recognize different types of queries, including those related to child exploitation or sexualization, and to not provide a response to certain queries which may be harmful or illegal, including child exploitative materials.
- **Continual Testing**: Dedicated teams of internal experts are testing our models through red teaming exercises. These teams work with internal child safety experts and use their institutional knowledge of child safety risks online to test our models with terms and prompts that may be used by those seeking to harm children, allowing us to identify and address inappropriate responses.
- **Removing Violating Content from Responses**: Building on our long-standing investment in technology that helps to proactively find and remove child exploitative content, we have implemented new technology into our models that works to prevent such content from responses before they are shared with people, in the event the model were to initially generate a response. For example, if someone prompts our AI to create content that could exploit or harm children, our proactive technology works to scan responses and prevent those that may relate to child exploitative content from being shown.
- **Providing Feedback on Responses**: We have developed feedback tools so people can flag responses that they perceive to be unsafe or offensive, and we will use this feedback to continue training the models and improve our ability to restrict our AI from providing such responses.

Addressing the challenge of deep fakes requires a whole-of-industry approach. That is one reason why we welcomed the White House's Voluntary Commitments on AI. Specifically, we will work with industry peers to align on technologies that can make it easier for us and other providers to detect when someone shares content that has been AI-generated. This approach will also pose challenges, as new companies creating AI tools will constantly emerge. Moreover, we know that bad actors will continue trying to find ways to circumvent our detection capabilities.

To that end, we continue to partner with the Partnership on AI, in the hope of developing common standards for identifying and labeling AI-generated content, as well as mitigating deceptive AI-generated content, across the industry. In particular, we support efforts to develop industry standards about how and when to apply watermarks to photorealistic images—and we think this is a place where Congress can help drive the consensus forward.

Further, as the difference between human and synthetic content gets blurred, we understand people want to know where the boundary lies. That is why we have been working with industry partners to align on common technical standards that signal when a piece of content has been created using AI. Being able to detect these signals will make it possible for us to label AI-generated images that people post to Facebook and Instagram. We are building this capability now, and in the coming months we will start applying labels in all languages supported by our apps.

Question 31. **What safeguards are you putting in place to ensure that bad actors do not leverage open source artificial intelligence to develop biological weapons?**

Please see the response to your Question 28.

Question 32. **In 2023, Meta's LLaMA model leaked online. Did Meta's General Counsel or any lawyer ever raise concerns with you about developing or releasing this open-source model?**

We believe that the responsible development of large language models depends on more researchers working on it, rather than developing technology in a silo. Consistent with this and in an effort to accelerate responsible development of Llama, we decided to release Llama in four different sizes to approved members of the AI research community in February of last year. Sharing Llama allowed other researchers to, for example, more easily test new approaches to limiting or eliminating model biases or toxicity in large language models. As a result, we have received—and continue to receive—valuable feedback on security and safety. For example, University of Edinburgh, Google Research, and Macquarie University researchers published research offering approaches to improve Llama's performance on generating accurate information.

The use of open source models promotes transparency and allows more people to access AI tools, democratizing this technology and decentralizing AI expertise, promoting innovation and driving more rapid progress in the industry. For more on our approach to open source models, please see the response to your Question 10.

Question 33. **What safeguards has Meta put in place to ensure that such models do not leak again in the future?**

Please see the response to your Questions 10, 28, and 32.

***Question 34*. Molly Russell was a 14-year-old girl who took her life in 2017 after viewing harmful content related to suicide, self-harm, and depression on Instagram. In a 2022 ruling, the inquest found that she "died from an act of self-harm while suffering from depression and the negative effects of online content." What steps did your company take to investigate the particular circumstances of Ms. Russell's exposure to negative content and death?**

In 2019, we undertook significant work to review our suicide and self-injury policies. The concerns raised by Molly Russell's family regarding suicide and self-injury content on Instagram and other platforms played an important part in that process. This review resulted in a number of developments to our approach including, for example, the extension of our suicide and self-injury policy to prohibit not only content that promotes suicide or self injury, but also any graphic suicide and self-injury imagery and real-time depictions of suicide or self-injury, irrespective of the context in which it is posted.

We continue to consult with experts—including the Suicide and Self Harm Expert Advisory Group—on a regular cadence and implement changes informed by their advice. For example, in January 2024, we updated our policies around age-inappropriate content to start restricting such content related to suicide and self harm from teens' experience on Instagram and Facebook. While we already aim not to recommend this type of content to teens in places like Reels and Explore, these changes mean we will aim to no longer show it to teens in Feed and Stories, even if it is shared by someone they follow.

Additionally, since 2006, we have worked with suicide prevention experts to support the Meta community. For those who may post potential suicide and self-harm content, we use proactive detection technology to send this content to our teams for prioritized review. When someone searches for, or posts, content related to suicide, self-harm, eating disorders, or body image issues, they will see a pop-up with tips and an easy way to connect to organizations like the National Alliance on Mental Illness (NAMI) in the US.

We also continue to improve our efforts to detect and remove policy-violating content. Using machine learning technology, we have expanded our ability to identify possible suicide or self-injury content, and, in many countries, we are able to use this technology to get timely help to people in need. This technology uses pattern-recognition signals, such as phrases and comments of concern, to identify possible distress. We also use artificial intelligence to prioritize the order our team reviews reported posts, videos, and livestreams. This helps us to efficiently enforce our policies and allows our reviewers to evaluate urgent posts and contact emergency services when members of our community might be at risk of harm.

To track our progress and demonstrate our continued commitment to making Facebook and Instagram safe, we regularly release the Community Standards Enforcement Report. This report shares metrics on how we are doing at preventing and taking action on content that goes against our Community Standards, including suicide and self-injury. For example, between October and December 2023, of the suicide or self-harm content we actioned, we proactively identified 99.1% of that actioned content, before it was reported to us on Instagram, and 99.4% on Facebook.

***Question 35*. What steps is your company taking to ensure that other children do not have similar experiences or suffer the same fate?**

Please see the response to your Question 34.

***Question 36*. Meta reported over 30 million cases of child sexual exploitation to NCMEC in 2023. What do you believe is the reason for such a high incidence of child sexual exploitation on your platforms?**

To understand how and why people share child exploitative content on Facebook and Instagram, we previously announced that we conducted an in-depth analysis of the illegal child exploitative content we reported to NCMEC. We found that more than 90% of this content was the same as or visually similar to previously reported content. And copies of just six videos were responsible for more than half of the child exploitative content we reported in that time period. While this data indicates that the number of pieces of content does not equal the number of victims, and that the same content, potentially slightly altered, is being shared repeatedly, one victim of this horrible crime is one too many.

The fact that only a few pieces of content were responsible for many reports suggests that a greater understanding of intent could help us prevent this revictimization. We worked with leading experts on child exploitation, including NCMEC, to develop a research-backed taxonomy to categorize a person's apparent intent in sharing this content. Based on this taxonomy, we evaluated 150 accounts that we reported to NCMEC for uploading child exploitative content, and we estimated that more than 75% of these people did not exhibit malicious intent (*i.e.*, did not intend to harm a child). Instead, they appeared to share for other reasons, such as outrage or in poor humor (*i.e.*, a child's genitals being bitten by an animal). Based on our findings, we developed targeted solutions.

We have spent more than a decade developing policies and technologies to help keep young people safe and to keep predators from attempting to use our services to connect with one another. Our comprehensive approach includes cutting-edge technology to prevent, detect, remove, and report violations of our policies that prohibit child exploitation, as well as providing resources and support to victims. We work with professionals, collaborate with industry, and

support law enforcement around the world to fight the online exploitation of children. For example, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations.

As such, we have built sophisticated technology so we can find, remove, and report more exploitative content than any other company that reports to NCMEC. Frequently, our systems *immediately* detect CSAM being uploaded, enabling us to thwart the attempt to distribute the content. Even in such instances, we submit a report, which reflects not incidences of child sexual exploitation on our platforms, but rather *failed* attempts by bad actors to use our platform for this abhorrent crime. Our automation has enabled us to report at high volumes because we detect at scale, which in turn drives the volume of reporting. And we will continue to refine our systems, and we call upon the rest of the industry to do the same.

NCMEC has acknowledged Meta continues to be an industry leader in this work and that Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs." We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

Question 37. **Instagram includes age-verification features, but does not require users to verify their age on account setup. Why not?**

We do require everyone to be at least 13 years old before they can create an Instagram account. Our community members cannot create an Instagram account without listing their birthdate, and we block people from repeatedly changing their birthdate. If someone enters a birthdate that puts them under the age of 13, they are unable to create an account. Creating an account with inaccurate, incomplete, or out-of-date information is a violation of our terms. This includes accounts registered on behalf of anyone under age 13. In the absence of updated laws that create requirements around how age is verified online, we have invested in technologies and tools that help verify age while protecting privacy. Our investments are working: in December 2022, we announced that since we had begun testing our age verification tools on Instagram in June of that year, we were able to stop 96% of the teens who attempted to edit their birthdate and age from under 18 to 18 or over on Instagram.

The difficulty in understanding someone's age online is not unique to Meta or even social media, and it warrants a simple solution that would apply across the industry. We support federal legislation that requires app stores to get parents' approval whenever their teens under 16 download apps. With this solution, when a teen wants to download an app, app stores would be required to notify their parents, much like when parents are notified if their teen attempts to make a purchase. Parents and guardians can decide if they want to approve the download.

Parents and guardians can also verify the age of their teen when setting up their phone, negating the need for everyone to verify their age multiple times across multiple apps.

This solution also helps to preserve privacy. By verifying a teen's age on the app store, individual apps would not be required to collect potentially sensitive identifying information. Apps would only need the age from the app store to confirm that teens are old enough to register for a platform and place them in the right experiences for their age group. Parents and teens will not need to provide hundreds of apps with information like government IDs. Instead they would provide it in just one place, the app store that comes with the device. In many cases, the app store already is collecting this information for its own purposes.

Question 38. **How many of Instagram's users are under the age of 13?**

Meta recognizes the need to keep people who are too young off Facebook and Instagram. Both Facebook's Terms of Service and Instagram's Terms of Use in the United States require people to be at least 13 years old to sign up for Facebook and Instagram. Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

Question 39. **Does Meta—for Facebook, WhatsApp, and Instagram—have access and provide copies of users' direct message communications to federal law enforcement upon request?**

In general, Meta responds to government requests for data in accordance with applicable law and our terms of service. Each and every request we receive is carefully reviewed for legal sufficiency and we may reject or require greater specificity on requests that appear overly broad or vague. With respect to US legal process, Meta Platforms, Inc. discloses account records solely in accordance with our terms of service and applicable law, including the federal Stored Communications Act ("SCA"), 18 U.S.C. Sections 2701-2712. If a chat is protected by end-to-end encryption, Meta will not be able to provide unencrypted message content as part of any response to any legal process that it receives unless the message has been reported to Meta. Meta will continue to provide message and call logs, as well as IP data.

We are also deeply committed to end-to-end encryption, which means only the sender and recipient can see the contents of a message. Meta does not have access to direct message

content where end-to-end encrypted on our services, except when users report that content to us. We are rolling out end-to-end encryption on direct messages by default on Messenger, and people have the option to initiate end-to-end encrypted direct messages in Instagram. Personal messages on WhatsApp are end-to-end encrypted. We believe end-to-end encryption is one of the strongest tools we have to protect the privacy and security of people and their messages, which include some of our most personal communication shared with friends and family.

Question 40. **As Meta continues developing the Metaverse, what do you think will be the greatest risks for children in this new space?**

Teens have become fans of popular virtual experiences across the industry—this makes it crucial that we build age-appropriate and positive experiences for them in virtual reality (VR). Doing so is core to our responsible innovation principles and our efforts to build safer experiences for young people.

Meta Horizon Worlds is one app among many multiplayer experiences in VR. Like other experiences available on our platform, Meta Horizon Worlds is a social experience in which people can explore, play, and create worlds with others. We have welcomed teens into Meta Horizon Worlds, where they can spend meaningful time with friends who may be separated by distance, explore their interests and passions, and express themselves. We have additional protections and tools in place to help provide age-appropriate experiences, including strong defaults for features like voice mode, Meta Horizon profile privacy settings, and active status settings, discussed in more detail below.

We know it is important for people to feel safe and in control of their experience and surroundings in Meta Horizon Worlds. To that end, we have built features into the app that help empower people, including:

- Pause, a space where you can take a moment away from other people and your surroundings.
- The ability to block and mute people.
- The ability to report people or content in real time.
- A personal boundary so other players cannot get too close, which makes it easier to avoid unwanted interactions.

We have supported teens with safety tools and built-in protections to facilitate an age-appropriate experience since opening up Meta Horizon Worlds to them, including:

- **Limiting interactions between teens + adults they do not know:** We want to help prevent teens from hearing from adults they may not know. That is why we take steps to help prevent interactions between adults and unconnected teens. For example, we do not display any adults a teen does not know in their "people you might know" list.
- **Meta Horizon profile privacy settings:** Teens are in control of who they follow and who can follow them back. Teens' profiles are automatically set to private, so they are able to approve or decline anyone who requests to follow them.
- **Active status settings:** By default, we do not show a teen's active status and Meta Horizon Worlds location to other people in Worlds. Teens are able to choose whether their connections can see if they are active and which public world or event they are in.
- **World and event content ratings:** We use content ratings to help ensure teens have age-appropriate experiences within Worlds. For example, mature world and event ratings prevent teens from finding, seeing, or entering spaces that contain mature content. Our policies prohibit teens from publishing mature worlds or events. Worlds violating this policy are removed.
- **Voice mode:** This feature transforms the voices of people a teen does not know into quiet, friendly sounds, giving teens more control over who can communicate with them. It will also garble the teen's voice, so people they do not know can not hear them. We turn garbled voices on automatically for all teens by default within voice mode.
- **Blurred chats:** This feature is turned on by default for teens and obscures messages from people the teen doesn't know in world chat, and the teen's messages appear blurred to them, too. Parents can also view, adjust, and lock this feature within Parental Supervision.
- **World chat filter:** This filter automatically hides words or phrases that might be upsetting or offensive in world chat. World chat filter is turned on by default for all people, including teens. Parents can also view, adjust, and lock this feature within Parental Supervision.

We also prohibit certain content that people using the app might find offensive. In addition to what is not allowed under the Code of Conduct for Virtual Experiences, we do not allow the following in Meta Horizon Worlds:

- Content that depicts or promotes the use of illegal drugs or abuse of prescription drugs.
- Content that promotes criminal or dangerous activity.
- Content that attempts to buy, sell or trade real life regulated goods, such as firearms, blades, alcohol and tobacco.

Additionally, everyone—including teens—can cast their experience from Meta Quest, allowing them to share what they're seeing in VR and Worlds with parents, guardians, or others around them.

Further, Meta works with over 500 women's safety NGOs around the world through regional roundtables and during the UN Commission on the Status of Women to get ongoing feedback on our safety tools and collaborate with civil society experts to make all online spaces welcoming and safe for everyone.

***Question 41*. Have you ever had any contact with Jeffrey Epstein? If so, what was the nature of the interaction(s)?**

Mark Zuckerberg met Mr. Epstein in passing one time at a dinner honoring scientists that was not organized by Mr. Epstein. He did not communicate with Mr. Epstein again following the dinner.

***Question 42*. How many times have human smugglers used your platforms to advertise and/or deliver their services (i.e., helping people cross illegally into the United States)?**

Human trafficking and exploitation is abhorrent and not allowed on our platform. In an effort to disrupt and prevent harm, we remove content that facilitates or coordinates the exploitation of humans, including human trafficking or human smuggling. We have long-standing policies and protocols to combat human exploitation, including, but not limited to, human smuggling. In May 2019, after consulting third-party experts, academics, and practitioners from around the world, Meta consolidated multiple related policies into a Human Exploitation Policy. The policy consolidation was consistent with the feedback Meta received from these individuals and groups, who encouraged Meta to address a broad range of harmful and exploitative activities through one comprehensive Human Exploitation Policy.

After extensive consultations with experts and stakeholders, we developed an approach that seeks to reduce opportunities for exploitation while allowing asylum seekers to access information to help them make informed choices and we continue to evolve our policy when appropriate to address changing and emerging trends. Under our Human Exploitation Policy, Meta has and will continue to forbid criminal organizations and other human smugglers from using our platforms to offer or facilitate their services. At the same time, our policy does allow people to discuss and seek legal migration, especially in the context of escaping conflict, oppression, or otherwise unsafe conditions, their right to seek asylum, and their desire to escape these dire situations. In addition to our Community Standards, we also require advertisers to follow our Advertising Standards, and people or businesses selling on Marketplace to follow our Commerce Policies, which also prohibit any form of human exploitation.

If a person attempts to post content on Facebook or Instagram seeking cross-border smuggling services, we will remove the content and offer resources that provide information about the risks of engaging with smugglers, the signs of potential exploitation, and ways to seek legal migration, including asylum. We still allow content that asks for or shares information about personal safety and how to leave a country or seek asylum through a legal process, as well as content condemning or raising awareness about human smuggling (i.e., news reporting, civil society campaigns, or personal stories).

WhatsApp, for its part, operates differently than social media. WhatsApp does not enable users to search or to discover other unconnected people or groups, and does not use algorithms to prioritize the delivery of private messages. We prohibit WhatsApp's service from being used to coordinate or facilitate human exploitation. In addition, WhatsApp cooperates with valid legal processes from law enforcement and complies with US law regarding designated organizations, including narcotics traffickers, their organizations, and operatives designated under the Foreign Narcotics Kingpin Designation Act, including by banning accounts.

The informational resources were developed in consultation with expert organizations including the International Organization for Migration and align with international law and human rights standards. We regularly engage with outside experts to help us craft policies that strike the right balance between supporting people fleeing violence and religious persecution while prohibiting human smuggling on our platforms. Through consultations with these experts, we decided to focus on providing resources about safe, legal migration options to people who most urgently need the support. We have worked with experts across academia, victim services, and law enforcement to develop support pages on our Help Center for harms such as sex trafficking, organ trafficking, labor trafficking, and human smuggling. We work to proactively share this information with people on our platforms who may need support. Human exploitation can only be tackled with strong dedicated efforts amongst policymakers, civil society, academia, law enforcement and companies—so we work closely with experts such as Tech Against Trafficking and Stop the Traffik, and support education initiatives.

On both Facebook and Instagram, Meta uses machine learning to proactively identify and take action on potential human exploitation content; either by removing it automatically from the app or, where feasible, escalating it to human reviewers to take appropriate action. We have teams across investigations, engineering, research, policy, and integrity who are dedicated to anti-trafficking efforts, and we have invested in technology to proactively detect content and behavior related to human exploitation. We are constantly evaluating ways to improve our enforcement, so we can most effectively find and remove content that violates our policies. We support our ability to detect violating content related to human exploitation through major investments by our technical and operational teams.

We work to implement countermeasures—both on our platforms and via our external partnerships—to stop actors and businesses from using our services to commit crimes, at all stages of the exploitation lifecycle. Although we have tools to combat recidivism, we do find that human smuggling organizations often try to return to our platforms. As a result, Meta uses a variety of tools to disrupt criminal organizations, including designation under our dangerous organizations policies, conducting human review, and employing a range of artificial intelligence and network disruptions. Meta relies on people and technology to remove this content, and works with NGOs and other stakeholders to combat ways our platforms may be used by those who want to harm people.

***Question 43*. What efforts has your company taken to ensure that your platforms are not used to facilitate human smuggling?**

For more information on our Human Exploitation Policy, please see the response to your Question 42.

Questions from Senator Lee

Question 1. **Internal emails sent to high-ranking executives at Facebook/Meta dating back to 2019 indicate that you had a clear understanding of the substantial negative effects your products have on minors. In 2021, Arturo Bejar—former senior engineer and product leader at Facebook from 2009 to 2015, independent consultant and industry expert for the Instagram Well-being team from 2019-2021, and Technical Advisor for the Facebook Oversight Board in 2022—sent you a series of emails explaining the "staggering levels of abuse that teens aged 13 to 15 were experiencing every week." Do you find it acceptable that as many as 21.8 percent of 13 to 15-year-olds were the target of bullying within the previous week? Or that 24.4 percent of 13 to 15-year-olds received unwanted advances during that same time period?**

Respectfully, we disagree with Mr. Bejar's allegations. We do not use the types of surveys cited above to measure policy violations or how effective we are at keeping people from seeing policy-violating content. Instead, we use surveys like this as one way to help us understand how users experience Instagram and to help us get a sense of how people feel about their experience on our services, and we do use these surveys to inform our safety and well-being efforts. People's responses are personal to themselves and subjective. We are committed to making the user experience as positive as possible.

At Meta, we take bullying and harassment seriously. Bullying and harassment present a unique challenge and are complex issues to address because context is critical. We work hard to enforce against this content while also equipping our community with tools to protect themselves in ways that work best for them. We rely on a combination of user reports and technology to find this type of policy-violating content and remove it. And in Q4 2023, of the bullying and harassment content we actioned, we proactively removed 95.3% of the actioned content on Instagram and 86.5% of the actioned content Facebook before it was reported to us. We do not tolerate this kind of behavior because it prevents people from feeling safe and respected on our apps. And we recognize that bullying and harassment can be more challenging for minors, which is why our policies are intended to provide heightened protection for users between the ages of 13 and 18.

To combat bullying, we have also built a number of sophisticated features and tools to support the people on our platform. For example, we have created comment warnings when people try to post potentially offensive comments. During our initial tests, we found that, about 50% of the time, people edited or deleted their comments based on these warnings. We also developed Restrict, a tool by which people can restrict someone from commenting on their account. Once enabled, comments from a restricted person will only be visible to that person. A person can choose to view the comment, approve the comment so everyone can see it, delete it, or ignore it. We developed Restrict specifically in response to feedback from teens, because they told us they

wanted a more subtle way to block bullies without them knowing they had been blocked. Additionally, because bullying can be very contextual and not always identifiable by reviewers or our systems, we also give teens the option to turn on Hidden Words for comments and Direct Messages. Once on, comments and Direct Messages containing emojis, words, or phrases selected by the teen will be hidden—any emojis, words or phrases, even those that would not be easily identifiable as bullying to a reviewer or classifier. We encourage people to use tools available in our Safety Center to help protect against such behavior. We also have a Bullying Prevention Hub, developed with the Yale Center for Emotional Intelligence, which is a resource for teens, parents, and educators seeking support for issues related to bullying and other conflicts.

Question 2. **The 2022 Thorn Report revealed similar troubling results to those identified by Arturo Bejar. Thorn reported that roughly 1 in 5 minors who used Meta's three most popular apps—Facebook, Instagram, and Messenger—experienced online sexual interactions. 1 in 6 minors had online sexual interactions with someone who the child thought to be an adult. Despite Instagram restricting messaging between adults older than 19 from messaging minors starting in March of 2021, nothing prevents an adult from posing as a minor while creating an account and messaging children, a problem that cannot be solved by app-store age verification. Other such loopholes exist. What more will you do to prevent minors from receiving any and all sexual messages on your platforms?**

Pretending to be someone else is an explicit violation of our policies. Impersonation is one way criminals attempt to gain the trust of victims, which is one reason why our policies prohibit it. We have invested heavily in strengthening our technology to aid in keeping fake accounts off our platforms—to help address the problem at the root and combat downstream harms, such as sextortion. We cooperate with law enforcement in this space and respond to lawful information requests in prosecutions of scammers.

We work to protect people by helping to prevent unwanted contact across our apps and in our messaging services, especially between adults and teens. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists. We do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We also restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos or repeated messages from people they do not know. We also announced that we plan to introduce stricter default message settings

for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts, and we do not recommend teen accounts to these accounts, or vice versa. We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Finally, we also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are

not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

Question 3. **I feel strongly about privacy and believe one of the best protections for privacy in this high-tech world is end-to-end encryption. However, we also know that a great deal of grooming and sharing of CSAM happens on encrypted systems. Does Meta allow juvenile accounts on its platforms to use encrypted messaging services? Why do juvenile accounts need to have their messages encrypted?**

Meta recognizes the need to keep people who are too young off Facebook and Instagram. Both Facebook's Terms of Service and Instagram's Terms of Use in the United States require people to be at least 13 years old to sign up for Facebook and Instagram. Meta blocks any individual from creating a Facebook or Instagram account if the individual indicates they are under the age of 13. If we receive reports a user may be underage, we will investigate. When there is reliable evidence an individual is under 13, we will disable the account, and provide the user the opportunity to verify their age. When there is reliable evidence an individual user is over 13, they will be permitted to remain on the service. In certain instances, accounts will remain active because the account has insufficient activity from which to assess the account holder to be violating our terms as an under 13 individual.

We want teens to have safe, age-appropriate experiences on our apps, including on our encrypted services. We do not believe moving to an encrypted messaging environment means sacrificing safety. It is already widely used by other large messaging services to protect people's private messages and provide people with the privacy and security they expect when messaging friends and family. That is why we will continue to support encryption. But encryption alone is not a complete solution for privacy, safety, and security, which is why we will continue building features to help keep people safe.

Our approach to safe encrypted experiences is focused on three key elements: (i) preventing potential harm in the first place; (ii) giving people ways to control their experience; and (ii) responding to violations of our policies quickly. This approach is detailed in our whitepaper, Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging.

To address the potential for harm, we have built tools and policies specifically to help young people manage interactions with adults. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists. We also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

In addition, in an end-to-end encrypted environment, we use machine learning to proactively detect accounts engaged in malicious patterns of behavior. Our machine learning technology will look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors.

We have also built more than 50 tools, resources, and features to help protect teens. Parental supervision tools are available globally on Facebook and Messenger. Parents using supervision tools will be prompted to approve or deny their teens' (under 16) requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request. Parents using supervision tools also receive reports when their teen blocks someone or reports something. This notice encourages teens to add their parents to supervise their teens' accounts as an extra layer of support. For more information on these tools as well as to review resources from experts, visit our Family Center: https://familycenter.meta.com/.

To help us respond to violations of our policies quickly, we encourage people to report messages to us in both encrypted and unencrypted services. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting potentially suspicious activity and educating people on how to take action. These notices help people avoid scams, spot impersonations, and flag accounts that have been exhibiting potentially suspicious behavior that attempt to connect to minors.

***Question 4*. When a concerning interaction involving a juvenile account is flagged by your platform, what do you do to notify parents?**

Parents and teens using supervision tools can receive reports when their teen blocks someone or reports something. We are committed to helping to build safe, healthy, and supportive digital communities, so in addition to encouraging reporting of policy-violating behavior and content, we encourage people to use tools available in our Safety Center to help protect against such behavior. We also have a Bullying Prevention Hub, which is a resource for teens, parents, and educators seeking support for issues related to bullying and other conflicts. It offers step-by-step guidance, including information on how to start important conversations about bullying. The educator section of the Bullying Prevention Hub includes information for educators about what

to do if their student is being bullied, what to do if their student is a bully, and tips on prevention planning in school.

Question 5. **When you see anyone ask a juvenile to move to encrypted message service, do you alert the parent of that request?**

If parents and teens use supervision tools, parents can receive reports when their teen blocks someone or reports something. For example, on Messenger, parents can view who can message their teen (only their friends, friends of friends, or no one) and see if their teen changes this setting. And on Instagram, parents can see which accounts their teen is following and which accounts are following their teen. We are continuously exploring new ways to actively defend against predatory behavior, including ongoing enhancement of our detection and removal systems, because this is a highly adversarial space where sophisticated predators are constantly evolving their tactics to avoid detection. We also restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages anyone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos or repeated messages from people they do not know. We also let teens know when an account exhibiting potentially suspicious behavior has attempted to follow them on Instagram, and encourage young people to be cautious.[29]

We also use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts, and we do not recommend teen accounts to these accounts, or vice versa. We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations, and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Messenger's parental supervision tools allow parents to see how their teen uses Messenger, including how much time they are spending on messaging and information about their teen's message settings. Parents can also: view and receive updates on their teen's Messenger contacts list, as well as their teen's privacy and safety settings; get notified if their teen reports someone

[29] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

(if the teen chooses to share that information); view who can message their teen (only their friends, friends of friends, or no one) and see if their teen changes this setting; and can view who can see their teen's Messenger stories and get notified if these settings change.

Question 6. **When you flag CSAM for NCMEC, are you also making parents aware of these interactions when a juvenile account is involved?**

No. Given the seriousness of such allegations, the proper authorities—rather than individuals with custodial responsibility—are best equipped to investigate, intervene, and confer with parents, as appropriate. Aside from practical considerations that could create a safety risk for the child (for example, the report may involve a custodial parent or another person living within a child's home), there are potential legal implications that preclude concurrently notifying parents when we make NCMEC reports. For example, such reports may violate the Stored Communications Act (under which there is an explicit exception for disclosure to NCMEC) and defamation law.

As a general matter, Electronic Service Providers are legally obligated to report apparent violations of laws related to child sexual abuse material they become aware of to NCMEC's CyberTipline. To do so, we submit electronic reports that contain the apparent child exploitative image(s). We endeavor to make our reports robust and include various types of information allowed by law in order to protect people and our services.

We are proud of the strong relationship we have developed with NCMEC and continue to report all CSAM found globally to NCMEC's CyberTipline across our family of apps, in compliance with US law. We have built systems and review processes to prioritize and appropriately action violating content and accounts and, when appropriate, report it to NCMEC or law enforcement.

As a result of our robust and leading efforts, we find and report far more content to NCMEC than any other service today. In 2022, we made over 26 million reports between Facebook and Instagram. To put that in perspective, the rest of the industry made approximately 6 million reports to NCMEC combined.

Question 7. **In June, 2023, the Wall Street Journal, in concert with Stanford University and the University of Massachusetts Amherst, revealed an investigation they conducted into Meta's child protections. According to the investigation, Meta's products "connect[] pedophiles and guide[] them to content sellers via recommendation systems that excel at linking those who share niche interests." The report states that "Meta has struggled with these efforts more than other platforms both because of weak enforcement and design features." Facebook moderators did not find a group to be in violation of your Community**

Standards even though the group went by the name "Incest." What specifically have you done to shut down these pedophile networks in their entirety before their inception?

Preventing child exploitation is one of the most important challenges facing our industry today. Online predators are determined criminals who use multiple apps and websites to target young people. They also test each platform's defenses, and they learn to quickly adapt. That is why now, as much as ever, we are working hard to stay ahead. In addition to developing technology that roots out predators, we hire specialists dedicated to online child safety and we share information with our industry peers and law enforcement.

Last year, we created a task force to address allegations about the effectiveness of our work in this area. As part of that work we reviewed existing policies; examined technology and enforcement systems we have in place; and made changes that strengthened our protections for young people, banned predators, and removed networks they use to connect with one another. Our child safety teams continue to work on additional measures, including with Facebook Groups. As a result of this work, we expanded the existing list of child safety related terms, phrases and emojis for our systems to find. We have many sources for these terms, including non-profits and experts in online safety, our specialist child safety teams who investigate predatory networks to understand the language they use, and our own technology, which finds misspellings or spelling variations of these terms.

We also use technology to find relationships between terms that we already know could be harmful or that break our rules and other terms used at the same time. These could be terms searched for in the same session as violating terms, or other hashtags used in a caption that contains a violating hashtag. We combined our systems so that as new terms are added to our central list, they will be actioned across Facebook and Instagram simultaneously. For example, we may send Instagram accounts, Facebook Groups, Pages, and Profiles to content reviewers, restrict these terms from producing results in Facebook and Instagram Search, and block hashtags that include these terms on Facebook and Instagram.

To keep our systems updated, we work with our specialist child safety teams and child safety professionals to help us understand evolutions in coded language and to identify new and evolving terms, phrases, slang, and emojis that could be used in an attempt to evade our detection systems and bypass our policies. We also work with these professionals and organizations to build various interventions, including but not limited to our search interventions, safety notices, and safety education campaigns. We have also worked with child safety researchers to conduct collaborative research to improve child safety protections on our platforms.

In addition, and as discussed in response to your Question 2, we use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts. Specifically, we work to ensure that teens are not recommended to potentially suspicious adult accounts, and potentially suspicious adult accounts are not recommended to anyone (including to teens or other potentially suspicious adult accounts). We identified and removed more than 90,000 accounts from August 1, 2023 to December 31, 2023 as a result of this method.

On Facebook, we are using this technology to help better find and address certain Groups and Pages. For example, Facebook Groups with a certain percentage of members that exhibit potentially suspicious behavior will not be suggested to others in places like Groups You Should Join. Additionally, Groups whose membership overlaps with other Groups that were removed for violating our child safety policies will not be shown in Search. As we reported in December 2023, since July 1, 2023, we removed more than 190,000 Groups from Search. From July 1, 2023 to December 31, 2023, we also reviewed and removed over 21,000 Facebook Groups that violate our child safety policies.

We also hire specialists with backgrounds in law enforcement and online child safety to find predatory networks and remove them. These specialists monitor evolving behaviors exhibited by these networks—such as new coded language—to not only remove them, but to inform the technology we use to proactively find them. Between 2020 and 2023, our teams disrupted 37 abusive networks and removed more than 200,000 accounts associated with those networks.

Finally, we also collaborate with industry on new programs, such as Lantern. Lantern is a program from the Tech Coalition that enables technology companies to share a variety of signals about accounts and behaviors that violate their child safety policies. Lantern participants can use this information to conduct investigations on their own platforms and take action.

Question 8. **Why should parents have any confidence that they will be able to keep their children safe from predators on your platform when groups like "Incest" were permitted to remain on Facebook until the problem was published by a major news outlet?**

We have spent over a decade investing heavily in sophisticated technology that helps us proactively find violating content and accounts of this kind and remove them. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in enqueuing certain content for human review. We recently improved our defenses against such predators through our task force last year, discussed above in response to your Question 7.

In addition, we have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

Question 9. **Your statement in the January 31, 2024 hearing that "the existing body of scientific work has not shown a causal link between using social media and young people having worse mental health outcomes" is particularly concerning and misleading. The same study you cited continues:**

> **"Research suggests the harms of social media use can include encouraging young people to engage in unhealthy social comparisons and displacing time that could be given to sleep, exercise, studying, or other activities. Social media's distracting power can work against an adolescent's ability to sustain attention, a skill necessary for academic success and emotional adjustment. Some young people can also develop a dysfunctional need to use online games, which is related to anxiety and depression. It is possible that dysfunctional social media use may pose a similar problem."**

Are Meta's child safety protocols driven by your belief that there is no "causal link" between social media and child mental health issues? Do you perform internal investigations into Meta's impact on child mental health? Please disclose the findings of those investigations.

The consensus study by the National Academies of Science discussed at the hearing notes that the study's "review of published literature did not support the conclusion that social media causes changes in adolescent health at the population level."[30] Instead, the consensus study highlights that "[c]ontrary to the current cultural narrative that social media is universally harmful to adolescents, the reality is more complicated. Social media can connect adolescents

[30]Consensus Study Report Highlights by National Academies' staff based on the Consensus Study Report Social Media and Adolescent Health (2023), https://nap.nationalacademies.org/resource/27396/Highlights_for_Social_Media_and_Adolescent_Health.pdf.

with their friends and family and can serve as a place of safety and support . . . and can also serve as an educational resource and help cultivate and expand hobbies, interests, and creative pursuits"[31]

With so much of our kids' lives spent on mobile devices, it is important to ask and think about any effects on teens—especially on mental health and well-being. This is a critical issue, and we take it seriously. We work hard to help teens have positive experiences on our apps—it is why we have developed more than 50 tools to support them and their parents. We also work in collaboration with leading experts to better understand issues around mental health and well-being and to make features that help enable meaningful social interactions. There is a growing body of research that suggests social media can play a positive role in teens' lives and provide support to those who may be struggling or are members of marginalized groups. For example, an April 2022 Pew study reported that 80% of teens surveyed felt that social media helped them stay more connected to what was going on in their friends' lives, and 67% felt that they have people on social media who can support them through tough times.[32]

These are societal issues that go beyond any one company. We do internal research to find out how we can best improve the experience for teens, and our research has informed product changes as well as new resources. Understanding how technology impacts lives, especially teens' lives, is an important part of what we do. We also think more research is needed to understand the bigger picture, and we are supporting that research. For example, it is why we supported more funding for research in these areas, like passage of the Children and Media Research Advancement Act, which provides funding to the National Institutes of Health (NIH) to study the impact of technology and media on the development of children and teens. Additionally, we recently announced a pilot program, in partnership with the Center for Open Science, designed to contribute to the public's scientific understanding of how different factors may or may not impact well-being and inform productive conversations about how to help people thrive.

Mental health is a complex issue, and the existing body of scientific work has not shown a causal link between using social media and young people having worse mental health outcomes. The National Academies of Sciences report you reference evaluated results from more than 300 studies and determined that the research "did not support the conclusion that social media causes changes in adolescent mental health at the population level." It also suggested that social media can provide significant positive benefits when young people use it to express themselves, explore, and connect with others. We will continue to monitor research in this area and remain vigilant against any emerging risks.

---

[31] *Id.*

[32] https://www.pewresearch.org/internet/2022/11/16/connection-creativity-and-drama-teen-life-on-social-media-in-2022/

Question 10. **On January 10, 2024, Meta announced that you will restrict teens of all ages from accessing material promoting eating disorders, suicidal ideation, or self-harm. However, minors 16 and 17 years old still have the option to view content deemed "sensitive," including sexually suggestive materials. While you do not permit sexually explicit materials on your platforms, you do permit sexually suggestive materials. If you already block sexually suggestive materials for users under the age of 16, why not block all minors from viewing sexually suggestive materials?**

We restrict the display of nudity or sexual activity because some people in our community may be sensitive to this type of content. Additionally, we default to removing sexual imagery to prevent the sharing of non-consensual or underage content. Restrictions on the display of sexual activity also apply to digitally created content unless it is posted for educational, humorous, or satirical purposes.

Our nudity policies have become more nuanced over time. We understand that nudity can be shared for a variety of reasons, including as a form of protest, to raise awareness about a cause, or for educational or medical reasons. Where such intent is clear, we make allowances for the content. For certain content, we include a warning label so that people are aware that the content may be sensitive. We also allow photographs of paintings, sculptures, and other art that depicts nude figures.

However, we do take steps to block sexually suggestive material that lacks expressive value. Meta bans all sexually suggestive materials in advertisements. Under our Advertising Policies, ads must not contain nudity, depictions of people in explicit or suggestive positions, or activities that are overly suggestive or sexually provocative.

We have worked closely with global experts to align on what types of content could be sensitive for younger teens versus older teens. We also take steps to protect teens from inappropriate sexual content. Our content recommendation controls—known as "Sensitive Content Control" on Instagram and "Reduce" on Facebook—make it more difficult for people to come across potentially sensitive content or accounts in places like Search and Explore. We automatically place teens into the most restrictive content control setting on Instagram and Facebook. In addition, for those parents using parental supervision tools, if a teen tries to change their Sensitive Content Control from "Less" to "Standard", their parent will receive a notification prompting them to approve or deny the request. Virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Controls globally and in the US are still on this setting a year later.

Question 11. **It appears that Meta is capable of curating an experience specifically for a juvenile account—could you also restrict access to encrypted messaging systems for juvenile accounts?**

Please see the response to your Question 3.

Question 12. **You are aware of the emerging crisis of sextortion on your platforms. What are you doing to ensure the safety of children from adults posing as children? How can you improve your battle against sextortion?**

Having a personal intimate image shared with others can be devastating, especially for young people. It can feel even worse when someone threatens to share that image if a person does not give more photos, sexual contact, or money—a crime in most jurisdictions, commonly referred to as sextortion.

At Meta, we take a multi-faceted approach to combat sextortion. These efforts include (i) strict policies against content or activity that sexually exploits or endangers children, including sextortion; (ii) human and machine detection and enforcement, including specialized teams focused on combating sextortion and automated rules that detect and action at scale accounts; (iii) proactive investigatory work, including targeted investigations and removal of violating accounts to disrupt networks of bad actors attempting to exploit or financially extort minors, and—when appropriate—reporting them to the National Center for Missing and Exploited Children (NCMEC); (iv) safeguards to help prevent suspicious adult accounts from finding or interacting with teens on our apps, including parental controls; and (v) provide education and awareness resources to those who may had their intimate images shared online. These efforts are described in more detail below.

We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. We work to prevent this content, as well as inappropriate interactions between young people and suspicious accounts attempting to take advantage of them. We also prohibit behavior that exploits people, including sharing or threatening to share someone's intimate images. In addition, impersonation is one way criminals gain the trust of their sextortion victims, which is one reason why our policies prohibit it. This helps address the problem at the root and prevent downstream harms, like sextortion. We have invested heavily in strengthening our technology to keep fake accounts off Facebook and Instagram and we cooperate with law enforcement and respond to lawful information requests in prosecutions of scammers.

We have specialized teams working on combating sextortion. These teams are constantly working to understand the unique combinations of on-platform behaviors used by criminals seeking to exploit our services. We build automation rules that allow us to detect and action—at scale and with high-precision—accounts committing financial sextortion. Our teams continue to work on new solutions to address sextortion industry-wide, including by developing new ways to identify people potentially engaging in sextortion and thwarting their efforts.

In addition, our dedicated teams investigate and remove these criminals and report them to authorities, including law enforcement and NCMEC, when appropriate. We work with partners, like NCMEC and the International Justice Mission, to help train law enforcement around the world to identify, investigate and respond to these types of cases. We have developed a streamlined online process through which we accept and review all legal requests from law enforcement. If we have reason to believe that a child is in immediate or imminent danger, we may proactively refer a case to local law enforcement (as well as report it to NCMEC) to help safeguard the child.

We also work to protect people from sextortion by preventing unwanted contact across our apps and in our messaging services, especially between adults and teens. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists, which can be used as a lever by people trying to sextort others. We also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We also restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also introduced stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens can not be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We have developed ways to help people control their own experience. For example, people can choose who can message them, and can block anyone they do not want to hear from. People can report nude or sexual photos or videos of themselves or threats to share these images or videos to

our apps or technologies to prevent them from being reshared. Our teams review reports 24/7 in more than 70 languages. We have articles in our Help Center that help people understand how to report this activity on Facebook,[33] Instagram,[34] and Messenger.[35]

Finally, we use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts.[36] Specifically, we work to ensure that teens are not recommended to potentially suspicious adult accounts, and potentially suspicious adult accounts are not recommended to anyone (including to teens or other potentially suspicious adult accounts). We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Because this is an industry-wide concern, we also direct people to various tools to use if people have nude or sexual photos or videos to help prevent them from being shared or reshared online. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Anyone seeking support and information related to sextortion can visit our education and awareness resources, including the Stop Sextortion resources, developed with Thorn. These resources include immediate actions parents and teens can take if users are experiencing sextortion, as well as expert tips for teens, parents and guardians, and information for parents on how to talk to their teens about intimate images. We also worked with Thorn and their NoFiltr brand to create and promote educational materials that reduce the shame and stigma surrounding intimate images, and empower teens to seek help and take back control if they have shared them or are experiencing sextortion.

---

[33] How do I report an abusive photo on Facebook? | Facebook Help Center
[34] How to Report Things | Instagram Help Center
[35] Reporting Conversations | Messenger Help Center (facebook.com)
[36] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

We also need Congress to pass legislation requiring operating-system level age verification requirements. That would allow services like Instagram to more quickly identify suspicious behavior, such as adults pretending to be minors, and remove them from the app entirely before they can even make contact with a teen—in addition to the work we have already been doing to prevent this contact. This also allows parents to oversee and approve their teen's online activity in one place. When a teen wants to download an app, app stores would be required to notify their parents. Where apps like ours offer age-appropriate features and settings, parents can help their teens use them. Until then, we require people to provide their age when signing up for accounts on our services, which helps us to provide teens with age-appropriate experiences.

For more information about our work combating sextortion and intimate image abuse, please see our dedicated page in Safety Center, linked here: https://about.meta.com/actions/safety/topics/bullying-harassment/ncii.

Question 13. **If a person finds an image of themselves on one of your platforms that was uploaded without that person's consent, what process does Meta employ to allow that person to have their images removed? What is the maximum amount of time a person might have to wait for Meta to respond and have their images removed? What does Meta do to ensure that image cannot be shared in the future?**

As discussed in response to your Question 12, we encourage people to report content they think breaks our rules, and we prompt teens to report at relevant moments, such as when they block someone. People can report nude or sexual photos or videos of themselves or threats to share these images or videos to our apps or technologies to prevent them from being reshared. Our teams review reports 24/7 in more than 70 languages. We have articles in our Help Center that help people understand how to report this activity on Facebook,[37] Instagram,[38] and Messenger.[39] We have also developed technology that identifies accounts exhibiting potentially suspicious behavior, and we review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations, and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

[37] How do I report an abusive photo on Facebook? | Facebook Help Center
[38] How to Report Things | Instagram Help Center
[39] Reporting Conversations | Messenger Help Center (facebook.com)

We also provide information to people about other programs, such as Take It Down and Stop NCII. These programs help people report this activity to other participating technology companies, to aid in preventing the images from being reshared. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

For users that have a nude or sexual photo or video that was taken when they were under 18 and are concerned it will be shared or reshared online, they can take steps to help prevent further circulation through Take It Down. Similar to the process on StopNCII.org for people over 18, Take It Down assigns a unique hash value (a numerical code) to a person's image or video privately and without the image or video ever leaving their device. Once they submit the hash value to NCMEC, companies like Meta can use those hashes to identify whether there are matches on their platform, review and take action to prevent violating content from being posted on our apps in the future.

For more information, please see the response to your Question 12.

Question 14. **In the months leading up to the November 2020 election, you and your wife donated nearly $500 million to various election-related entities. These "relief grants" have been characterized by the New York Times as private intrusion into public election offices, stating [voters] "[m]ight consider the intervention of info-tech billionaires in the 2020 election to be a larger potential threat to our democracy" than the January 6, 2021, protest.**

Of the hundreds of millions of dollars you invested into swing states, 9 out of every 10 dollars in Pennsylvania went to counties that voted for Biden, 8 out of 10 dollars in Wisconsin went to Biden cities, and Biden counties in Georgia received nearly four times more money from your investment than Trump counties. Was it your intent to use your vast fortune to influence the outcome of the election?

No, it was not. The donations made by Mark Zuckerberg and Priscilla Chan to help support voting in 2020, during the unprecedented conditions of the pandemic, were not an initiative by the company. The private donations—made to non-partisan, non-profit organizations that distributed funds throughout the country—were apolitical and intended only to help ensure that all Americans could vote safely and have their vote counted.

Question 15. **The New York Post wrote: "Funding and managing elections has always been a government function, not a private one, and for good reason. Private organizations are not subject to the rules for public employees and institutions—they are not required to hold public hearings, cannot be monitored via open-records requests and other mechanisms of administrative and financial transparency, are not subject to the normal checks and balances of the governmental process and are not accountable to voters if the public disapproves of their actions." Do you intend on investing more money—personally or through any one of your foundations or corporations—towards influencing the outcome of the 2024 United States elections in any way?**

For additional information, please see the response to your Question 14.

Question 16. **You hired David Plouffe just after the 2016 election. What exactly did you hire him to do? Did he help you come up with a plan to help ensure the 2020 election did not go the same direction as the 2016 election?**

David Plouffe worked for the Chan Zuckerberg Initiative, a philanthropic organization and a separate entity from Meta. For more information about the Chan Zuckerberg Initiative, please visit czi.org.

Question 17. **In October 2020, three weeks before the presidential election, Facebook and Instagram suppressed news of the Hunter Biden laptop. You stated that "the distribution [of material pertaining to the laptop] was decreased," and you falsely labeled the laptop as "disinformation." A 2022 Technometrica Institute of Policy and Politics survey indicated that 47 percent of voters—including 71 percent of Democrats—would have changed their voting decision if they knew the contents of the laptop were real and not "disinformation" prior to the election. Do you consider your suppression of this vital information as election meddling?**

Regarding content about the October 14, 2020 *New York Post* story, given what happened in the 2016 election, we were concerned about potential election interference in the 2020 election. To be clear, at no point did we take any action to block or remove the content from our services. This reporting was always available on our services and people could, and did, engage with it. However, given the concerns raised, we took steps to slow the spread of content and provide fact-checkers the opportunity to assess it. After seven days, we lifted the temporary demotion on this content because it was not rated false by an independent fact-checker.

Question 18. **What specific instructions were you given by the FBI, the Department of Justice, or other government agencies regarding the Hunter Biden laptop story in the weeks leading up to the 2020 election? Include the names of the FBI and DOJ officials involved.**

We took independent steps, consistent with our policies to provide fact-checkers the opportunity to assess the content. After seven days, we lifted the temporary demotion on this content because it was not rated false by an independent fact-checker.

Question 19. **What specific steps did Facebook and Instagram take regarding the Hunter Biden laptop story to suppress its dissemination? What labels or barriers were placed on posts discussing the laptop?**

Please see the response to your Questions 17 and 18.

Question 20. **In 2021, you boasted about removing 18 million posts with COVID "misinformation." You permanently removed countless individuals from your platforms—including doctors and renowned immunologists—for sharing their opinions on the virus, its origins, and vaccinations. You were in regular communication with the White House and the CDC, and you accepted their demands that you censor certain ideas from your platforms. Accounts that shared the hypothesis that COVID-19 may have originated in the Wuhan Institute of Virology "could have led to a ban from the site entirely," which happened to China Scholar and New York Post contributor Steven Mosher in 2020. As of July, 2023, Mosher's account had yet to be reinstated. Despite this censorial stance, Facebook reversed its policy prohibiting speech about the lab-leak hypothesis in July of 2021. How many accounts did you ban for COVID "misinformation"? How many accounts remain banned? How many accounts were throttled or suppressed for their COVID content?**

We partnered with government agencies throughout the pandemic to connect people to authoritative health information and helpful resources, and we were transparent about the fact that we did so. In developing the standard for imminent physical harm as it relates to COVID-19, we consulted the CDC and other governmental health experts to assess whether a false claim, if believed by an individual, would increase the likelihood that the individual would contract or spread the virus. We updated the claims that we removed based on guidance from health authorities. For other false claims related to COVID-19, we have leveraged our third-party fact-checking program to reduce the distribution of false and misleading content. For example, in May 2021, Facebook stopped removing claims that COVID-19 was man-made, in response to a change in rating from third-party fact checkers.

Importantly, Meta's COVID-19 misinformation policies evolved alongside scientific research throughout the pandemic, and we stopped removing claims that the CDC and other health experts informed us were no longer harmful. We also reassessed whether our policies should remain in place altogether as the threat of COVID-19 subsided, vaccines became more available,

and scientific research regarding the pandemic improved. For example, in July 2022, Meta asked its Oversight Board for advice on whether our measures to address dangerous COVID-19 misinformation, introduced in extraordinary circumstances at the onset of the pandemic, should remain in place. The Board advised that we should stop removing those claims in countries that were no longer experiencing a state of emergency from COVID-19. Based on the Board's advice, we now take a more tailored approach to our COVID-19 misinformation rules consistent with the Board's guidance and our existing policies—our COVID-19 misinformation rules are no longer in effect globally, as the global public health emergency declaration that triggered those rules has been lifted, and we only enforce those specific policies in the few countries still having a COVID-19 public health emergency declaration in place, which the United States does not. We have also narrowed the claims enforced in those countries to only those that are prevalent on our platforms.

Question 21. **Facebook and Instagram accounts were routinely banned for stating that the COVID-19 vaccinations potentially lead to inflammation of the heart and surrounding tissue. The CDC currently reports that myocarditis and pericarditis are known side effects of the Pfizer and Moderna COVID vaccines. Moderna states that "[m]yocarditis . . . and pericarditis . . . have occurred in some people who have received mRNA COVID-19 vaccines . . . most commonly in males 18 years through 24 years of age." How many accounts remain banned for speaking about the risks of mRNA vaccines?**

Please see the response to your Question 20.

Question 22. **It has been reported that Meta will soon be partnering with the firm Logically.ai to monitor, and potentially automatically suppress content shared on your platforms that the software flags. Will you use Logically.ai's software to target conservative speech? What are the parameters you will set for this software—or similar software—to ensure that it does not interfere with future elections? What safeguards will you use to ensure otherwise protected speech remains unencumbered?**

No. We partner with Logically Facts, a distinct fact-checking business unit, in the UK only. For more on Logically Facts, please see here. More broadly speaking, we partner with nearly 100 independent fact-checking organizations around the world that review and rate viral misinformation in more than 60 languages globally. All our fact-checking partners are certified by the nonpartisan International Fact-Checking Network (IFCN).

Questions from Senator Padilla

Question 1. **In recent years, more companies in the tech sector are offering tools to enable caregivers to have a dialogue with minors in their care about healthy and safe internet activity. An important element in understanding whether these tools are helpful is understanding whether or not these tools are being adopted.**

a. **How many minors are on Instagram? And how many have caregivers that have employed your family center tools?**
b. **How many minors use Facebook and Messenger? And how many have caregivers that have employed your family center tools?**
c. **How many minors are on Meta Quest and Horizon? And how many have caregivers that have employed your family center tools?**
d. **How are you ensuring that young people and their caregivers are aware of these tools?**
e. **How are you ensuring that these tools are helpful to both minors and their caregivers?**

When we look at self-reported ages of our US daily active users, about 6% of Instagram accounts belong to teens under 18, and 1% of Facebook accounts belong to teens under 18. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. We educate parents in a variety of ways about our parental control features, including through Family Center's Education Hub, advertising campaigns, in-app promotion, and events with parents. Our ads encouraging parents to use our youth well-being tools and features were seen more than one billion times by tens of million people in the United States since August 2022.

We also work closely with external groups such as ConnectSafely and Net Family News to develop resources for parents and guardians to help them have meaningful and open conversations with their teens about being online. And through our partnership with Smart Design, we conducted co-design sessions with teens and parents and consulted with experts in the US, the UK, Ireland, Brazil, Japan and India. That co-design work invites young people, parents and experts to participate as collaborators in our design process, empowering them to provide input about how our services can meet their needs.

We built a Family Center to help teens and families build healthy online habits. The Family Center is a central place for parents and guardians to access supervision tools and resources from leading experts. It includes an education hub where parents and guardians can access resources from experts and review articles, videos, and tips on topics like how to talk to teens about social media. Parents and guardians can also watch video tutorials on how to use these new supervision

tools. Our vision for the Family Center is to eventually allow parents and guardians to help their teens manage experiences across Meta technologies, all from one place.

Among US teens adopting time management features on Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians & teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

Question 2. **In addition to keeping parents informed about the nature of various internet services, there's a lot more we need to do to inform our young people about unsafe, criminal conduct that is facilitated online. While many companies offer a broad range of "user empowerment" tools, it's helpful for us to understand whether young people even find these tools helpful or are actually adopting them. Meta shared that the company has developed over 30 tools to support teens and their parents.**

a. **For example, the company offers a comment warning when someone tries to post something offensive. How is this impacting user behavior?**
b. **The company launched "Take a Break." What is the rate of adoption amongst users and how many users have reported this to be a helpful tool in managing their time?**
c. **Can you describe how the company ensures that the tools it launches to help users are actually useful and impactful?**
d. **How are you ensuring that the burden is not on young people to make adult-level decisions about safety on the services that you operate?**
e. **Over the last 4 years, how often have you blocked products from launching because they were not safe enough for children, or withdrawn products from the market after receiving feedback on the harms they were causing?**

Safety and integrity are key to the experiences people have on our services, and Meta builds its services and continually updates them with safety and integrity in mind. We embed teams focusing specifically on safety and security directly into product development teams, allowing us to address issues during product development. And we offer integrity tools, built centrally, to individual product teams to allow them to build in preventative safeguards at the start. After products launch, we continue to monitor their impact, including by looking at integrity metrics, to help best serve our community.

With respect to the tools, features and resources you reference, a large majority of teens keep their default settings. For example, among US teens adopting time management features on

Instagram (Daily Limit, Take a Break, Quiet Mode), a large majority still use these features 30 days after initial adoption (over 90%, 80%, 70%, respectively). And virtually all (99%) teens defaulted into the "less" setting on Sensitive Content Control globally and in the US are still on this setting a year later. And over 90% of parents and teens in the US who use Instagram or Facebook supervision tools continue to retain supervision 30 days after initial adoption. And over 90% of guardians & teens in the US who choose Instagram or Facebook Supervision still use supervision 30 days after initial adoption.

Additionally, we show a Comment Warning on Instagram when someone tries to post a potentially offensive comment reminding them of our Community Guidelines and warning them that we may remove or hide their comment if they proceed. We have found these warnings really discourage people from posting something hurtful. For example, in a one-week period, we showed warnings about a million times per day on average to people when they were making comments that were potentially offensive. Of these, about 50% of the time the comment was edited or deleted by the poster based on these warnings.

Finally, we have policies, default settings, and tools in place designed to provide an age-appropriate experience for teens on our platform. Parental supervision tools are available globally on Facebook, Instagram, Messenger, and Horizon Worlds. Parents can use these tools, after being granted access by their teen, to see their teen's time spent, schedule breaks for their teens, see who their teens follow and who follows their teen. And our Family Center education hub provides parents with expert resources on supporting their teens' online. Parents of teens under 16 who use supervision tools are prompted to approve or deny their teens' requests to change their default safety and privacy settings to a less strict state—rather than just being notified of the change. For example, if a teen using supervision tries to change their account from private to public, change their Sensitive Content Control from "Less" to "Standard," or tries to change their direct message settings to hear from people they are not already following or connected to, their parent will receive a notification prompting them to approve or deny the request.

Question 3. **Existing detection tools for keeping child sexual abuse material from spreading online rely on hashed images of already identified CSAM imagery. There are tools like PhotoDNA and Google's CSAI match tool available for identifying this content. A challenge I hear raised frequently is identifying and removing novel images that have not already been hashed.**

a. **What would it take to develop better technology to accurately identify and limit the spread of novel CSAM images?**
b. **Are there interventions from Congress that could facilitate identification of CSAM?**

c. Based on your company's experience trying to address online sexual exploitation and abuse of minors, are there areas where Congress could be helpful in tackling this problem?

Over the years, we have invested heavily in sophisticated technology that helps us proactively find violating content and accounts of this kind and remove them. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in enqueuing certain content for human review.

We use AI and machine learning to proactively detect and take action against child nudity and previously unknown child exploitation and sexualizing content. As mentioned in your question, we use PhotoDNA and other photo- and video-matching technologies that detect identical or near-identical photos and videos of known child exploitative content, and we use Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess.

We also use technology to detect and remove Instagram Reels and Stories that violate our Community Guidelines, including by scanning for CSAM terms and for CSE indicators. In Q4 2023, we removed 16.2 million pieces of child sexual exploitation content on Facebook and 2.1 million pieces on Instagram. In Q4 2023, of the child sexual exploitation content we actioned, we detected 99% on Facebook and 95% on Instagram before it was reported by our users. In Q4 2023, Facebook and Instagram sent over 6 million NCMEC Cybertip Reports for child sexual exploitation. Of these reports, over 100,000 involved inappropriate interactions with children. Over 5.9 million reports related to shared or re-shared photos and videos that contain CSAM.

Our systems would be enhanced by federal legislation that would make it simpler for parents to oversee their teens' online lives, including legislation that would require app stores to get parents' approval when teens under 16 download an app. According to a recent Morning Consult poll,[40] parents across both sides of the aisle overwhelmingly support this approach. 81% of Democratic-leaning and 79% of Republican-leaning parents back federal legislation for parental approval of teen app downloads. Such legislation would help enable us to better know the age of people who use our platforms and would help improve our ability to ensure an age-appropriate experience. Additionally, federal legislation that would allow companies to retain CSAM for the limited purposes of machine learning training to prevent CSAM and legislation that would require law enforcement to provide feedback on NCMEC reports would be helpful in this area.

***Question 4*. AI models are making it easier to develop synthetic CSAM. These are either altered images of real people, or wholly synthetic individuals. Policymakers are grappling**

---

[40] https://pro-assets.morningconsult.com/wp-uploads/2023/12/US-Parents-Study-on-Teen-App-Downloads_Memo.pdf.

with what this will mean for law enforcement efforts to hold perpetrators accountable and identify children who are being harmed. In addition to processing a higher volume of Cybertips, investigators will have the added challenge of determining whether the victim in the scenario is in fact a real person. And cases are already being reported where AI generative technologies are being employed to facilitate the grooming and sextortion of minor victims.

a. **What are you doing to identify and remove AI-generated CSAM on your services?**
b. **Do you flag for NCMEC if you perceive the CSAM to be AI-generated?**
c. **How prevalent is this kind of content?**
d. **How do you anticipate the rise of AI-generated CSAM will impact NCMEC's ability to process and refer Cybertips to law enforcement?**

Child exploitation is a horrific crime that we work aggressively to fight on and off our services. We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. Our policies prohibit the sexualization of minors and child sexual abuse material—whether it is AI-generated or not—and we proactively work to find and remove this content. Our policies and enforcement are designed to adapt in this highly adversarial space, and we are actively monitoring new trends in AI-generated content.

We have processes in place to remove policy-violating content, regardless of the context or the person's motivation for sharing it. We also have developed aggressive, cutting-edge technology to help prevent, find, and remove policy violating content. Additionally, when we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children (NCMEC), in compliance with applicable law. In addition to this technology, we have invested in specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to both find and review potentially violating content, accounts and adversarial networks.

For example, we promptly disable accounts for various violations of our child exploitation policies, such as the apparent malicious[41] distribution of CSAM or sexual solicitation of children. We also recognize that predators may attempt to set up multiple accounts to evade enforcement of our policies. That is why when we disable accounts for these severe violations, we also work to disable explicitly linked accounts (where a person has linked their Facebook and Instagram profiles), high-confidence linked accounts (where we have high confidence that the same person

[41] We distinguish between "malicious" and "nonmalicious". In the "malicious" group are people we believe intended to harm children with their content, and in the "nonmalicious" group are people we believe, based on contextual clues and other behaviors, likely did not intend to cause harm to children. (For example, they shared the content with an expression of abhorrence.) Regardless, and consistent with US law, these actors are reported to NCMEC.

is using multiple accounts), and restrict those devices from setting up future accounts. We also utilize technology and teams to detect and eliminate abusive networks to take on predators who attempt to use our services to connect online.

We work to minimize the possibility of illegal child sexual abuse material being used to train our AI models. We also work with experts and industry partners to help prevent Generative AI models from being used to harm children, and we routinely test and retrain our models to help our AI features provide experiences that are safer and more helpful for young people.

We also strive to use a number of protections in our generative AI, including:

- **Training Our Model to Recognize Exploitative Queries**: We are training our models to recognize different types of queries, including those related to child exploitation or sexualization, and to not provide a response to certain queries which may be harmful or illegal, including child exploitative materials.
- **Continual Testing**: Dedicated teams work with internal child safety experts and use our institutional knowledge of child safety risks online to test our models with terms and prompts that may be used by those seeking to harm children, allowing us to identify and address inappropriate responses.
- **Removing Violating Content from Responses**: Building on our long-standing investment in technology that helps to proactively find and remove child exploitative content, we have implemented new technology into our models that works to prevent such content from responses before they are shared with people, in the event the model were to initially generate a response. For example, if someone prompts our AI to create content that could exploit or harm children, our proactive technology works to scan responses and prevent those that may relate to child exploitative content from being shown.
- **Providing Feedback on Responses**: We have developed feedback tools so people can flag responses that they perceive to be unsafe or offensive, and we will use this feedback to continue training the models and improve our ability to restrict our AIs from providing such responses.

Finally, we are proud of the relationship we have developed with NCMEC and continue to report all apparent CSAM found globally to NCMEC's CyberTipline across our family of apps. We have developed sophisticated technology to proactively seek out this content, and as a result we find and report more CSAM to NCMEC than any other service today. We also continue to encourage user reporting, as it is important to provide helpful context to take action against people who violate our Community Standards and an opportunity to support victims. Additionally, Instagram and Facebook are founding members of Take It Down—a service by

NCMEC designed to proactively prevent young people's images, including AI-generated content, from spreading online.

- **Recently, A.I.–generated explicit images of a major pop superstar were distributed widely online without her consent. That story drew attention to a growing problem over the last year facilitated by AI tools: the generation of deepfake, nonconsensual, sexually explicit imagery of everyday people, including our young people. Will you commit to reporting on the prevalence of this new problem and the steps your company is taking to address this horrendous abuse?**

We publish Community Standards and Community Guidelines governing the types of content and behaviors that are acceptable on Facebook and Instagram. These policies apply to all content on our platforms, including content generated by AI. We prohibit pornography and sexually explicit content on our platforms, as well as offers or asks for pornographic material (including, but not limited to, sharing of links to external pornographic websites). We further prohibit apps that offer to create this type of imagery. We remove images that depict incidents of sexual violence and intimate images shared without the consent of the person(s) pictured. We also prohibit derogatory sexualized manipulated imagery of real people. When we find this content, we work to remove it, regardless of how it is created.

Meta has dedicated significant resources to detecting content on our platform, including AI-generated content, that violates our policies. Our investments have allowed us to build technologies to proactively identify content, prioritize the most critical content for review, and act on content that violates our policies. We enforce our policies through a combination of people and technology that work to identify violations of our Community Standards across the billions of pieces of content that are posted to our platform every day. For example, our systems flag content that may violate our policies, people who use our apps report content to us they believe is questionable, and our own teams review content. We remove content that violates our policies quickly and at scale, with the help of media-matching technology that we built to find content that is identical or near-identical to photos, videos, text, and even audio that we have already removed. We have also built a parallel content review system to flag posts that may be going viral—no matter what type of content it is—as an additional safety net. This helps us catch content that our traditional systems may not pick up. We use this tool to detect and review Facebook and Instagram posts that were likely to go viral and take action if that content violated our policies.

Addressing the challenge of deep fakes requires a whole-of-industry approach. That is one reason why we welcomed the White House's Voluntary Commitments on AI. Specifically, we will work with industry peers to align on technologies that can make it easier for us and other providers to detect when someone shares content that has been AI-generated. This approach will also pose challenges, as new companies creating AI tools will constantly emerge. Moreover, we

know that bad actors will continue trying to find ways to circumvent our detection capabilities. To that end, we continue to partner with the Partnership on AI, in the hope of developing common standards for identifying and labeling AI-generated content, as well as mitigating deceptive AI-generated content, across the industry. In particular, we support efforts to develop industry standards about how and when to apply watermarks to photorealistic images—and we think this is a place where Congress can help drive the consensus forward.

Further, as the difference between human and synthetic content gets blurred, we understand people want to know where the boundary lies. That is why we have been working with industry partners to align on common technical standards that signal when a piece of content has been created using AI. Being able to detect these signals will make it possible for us to label AI-generated images that people post to Facebook and Instagram. We are building this capability now, and in the coming months we will start applying labels in all languages supported by our apps.

For more information, please see our responses to your Questions 4(a)-(d).

- **Are there technical or legal barriers that your company has identified preventing thorough redteaming of AI models to ensure they do not generate CSAM?**

Existing federal law makes red teaming exercises in this space more difficult, as it currently provides no immunity for such efforts. Nonetheless, we are working within the bounds of these laws to help ensure that our testing is as extensive as legally permissible.

We do not allow content or activity that exploits or harms children across Meta technologies, including in generative AI. We are working with experts and partners in the technology industry to help prevent generative AI services from being used to harm children, including through red teaming exercises that test our models, and we are routinely testing and retraining our models to help ensure that our AI features provide experiences that are safe and helpful for young people.

For more information, please see our responses to your Questions 4(a)-(d).

Question 5. **How companies choose to allocate their resources illustrates their true priorities.**

a. **What percentage of your company's budget is dedicated to addressing child safety on your platform?**

We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone.

b. **What process or assessment of risk on the platform informed that figure?**

For more than a decade, we have invested in teams and technology to combat child exploitation online, and these teams continuously explore new ways to defend against predatory behavior, including by adapting and expanding our detection systems to find and remove accounts that violate our child safety policies. We have specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to review potentially violating content and report findings to the NCMEC. In addition to enforcement actions, these teams closely collaborate with other teams internally about trends and signals to enhance our systems, tools, and policies to account for new trends and adversarial behaviors and protect minors from harm. We take the issues of safety and well-being on our platforms very seriously, especially for the youngest people who use our services. We are committed to working with parents and families, as well as experts in child development, online safety, and children's health and media, to ensure we are building better services for families. We are always working on new tools, re-evaluating our policies, and continually investing in detection technology to ensure we are proactively tackling the problem as best we can, as we know how important it is to get child safety right.

c. How many layers of leadership separates your trust and safety leaders from you?

Meta's Global Head of Safety works closely with Meta's executive team. The work of this global team is core to our mission of designing and building services that bring people together. This global team, responsible for ensuring that Meta remains a leader in online safety, works tirelessly with colleagues across the company to put in place the right policies, services, and precautions so that the people who use our services have a safe and positive experience. The Global Head of Safety also coordinates efforts of Meta's Safety Advisory Council, a team of leading safety organizations from around the world who provide Meta with cutting-edge research and advice on best practices, particularly relating to young people and other vulnerable groups.

Question 6. **The companies represented at the hearing have the money and resources to hire teams of Trust & Safety professionals and build bespoke tools to aid with content moderation and integrity work as well as the detection of content like CSAM on their services. This is not necessarily the case for the rest of the tech sector. These are industry-wide problems and will demand industry-wide professionalization and work.**

a. What is Meta currently doing to support access to open-source trust & safety tools for the broader tech ecosystem?

Child protection requires a global and comprehensive response from industry, law enforcement, government, civil society, and families, which is why we are committed to working with child-safety stakeholders to build and support the child-safety ecosystem. We build technology specifically to help tackle some of the most serious online risks, and we share it to help our whole industry get better.

Since 2019, we have also made two technologies—PDQ and TMK-PDQF—publicly available which detect identical and nearly identical photos and videos. We use PhotoDNA and other photo- and video-matching technologies that detect identical or near-identical photos and videos of known child exploitative content, and we use Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess. We also use technology to detect and remove Instagram Reels and Stories that violate our Community Guidelines, including by scanning for CSAM terms and for CSE indicators.

Our collaborative work to address child safety does not stop with improving our own services. We also are deeply committed to improving the entire ecosystem and have engaged with child safety nonprofits and academic researchers to complete child safety research with fieldwide impact. Our efforts with these professionals also include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, among other things. Additionally, to help safety stakeholders identify and respond to the most high priority reports at NCMEC, we funded and helped rebuild their case management tool to ensure investigators can get to the most important cases quickly.

Because this is an industry-wide concern, we also direct people to various tools to use if people have nude or sexual photos or videos to help prevent them from being shared or reshared online. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

On Take It Down, young people or their guardians can submit a case to proactively search for attempted uploads of their intimate images on participating platforms. Take It Down allows people to only submit a hash—rather than the intimate image or video itself—to NCMEC. Hashing turns images or videos into a coded form that can no longer be viewed, producing hashes that are secure digital fingerprints. Once a person submits the hash to NCMEC, companies like ours can use those hashes to find any copies of the image, evaluate any matches of images attempting to be uploaded to confirm they violate our policies, block the upload, and help prevent the content from being posted on our platforms in the future—helping to return power and control back to the victim.

Anyone seeking support and information related to sextortion can visit our education and awareness resources, including the Stop Sextortion resources we developed with Thorn. These resources include immediate actions parents and teens can take if users are experiencing sextortion, as well as expert tips for teens, parents and guardians, and information for parents on

how to talk to their teens about intimate images. We also worked with Thorn and their NoFiltr brand to create and promote educational materials that reduce the shame and stigma surrounding the sharing of intimate images, and empower teens to seek help and take back control if they have shared them or are experiencing sextortion.

Additionally, we collaborate across the industry through organizations like the Technology Coalition, an industry association dedicated solely to eradicating the sexual exploitation of children online. In 2020, Meta joined Google, Microsoft, and 15 other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual abuse. This project includes a renewed commitment and investment from the Technology Coalition, expanding its scope and impact to protect kids online and help guide its future work. Project Protect focuses on five key areas: tech innovation, collective action, independent research, information and knowledge sharing, and transparency and accountability. We also announced our recent participation in Lantern, a Tech Coalition program that enables technology companies to share a variety of signals about accounts and behaviors that violate their child safety policies. Lantern participants can use this information to conduct investigations on their own platforms and take action. Meta was a founding member of Lantern, providing the Tech Coalition with the technical infrastructure that sits behind the program and encouraging our industry partners to use it. We manage and oversee the technology with the Tech Coalition, ensuring it is simple to use and provides our partners with the information they need to track down potential predators on their own platforms.

Question 7. **One necessary element of keeping our kids safe is preventing harms in the first place. The National Center for Missing and Exploited Children partnered with the White House, the Department of Justice, and the Department of Homeland Security to create "The Safety Pledge" initiative to combat online child exploitation in September 2020. I understand more government backed public awareness campaigns are being developed.**

a. **Are you partnering with the federal government to distribute health and safety resources to young people?**

We have spent more than a decade working on these issues and have developed more than 50 tools, features and resources to support teens and their parents. We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone. We regularly consult with experts in adolescent development, psychology, and mental health to make our platforms safe and age-appropriate for young people, including improving our understanding of which types of content may be less appropriate for young people. Our teams also work with various outside stakeholders, including in government, to receive input on things that are happening on our platforms. For example, as we recently announced, we are partnering with the US Department of Homeland Security to launch Know2Protect, a national campaign to raise awareness about online child sexual

exploitation. The campaign—which will educate and empower people on ways to prevent and combat child safety risks both on and offline, explain how to report, and offer resources for survivors—aligns with our industry-leading efforts to prevent child safety harms before they happen.

Anyone seeking support and information can visit our education and awareness resources, including the Stop Sextortion resources, developed with Thorn. These resources include immediate actions parents and teens can take if users are experiencing sextortion, as well as expert tips for teens, parents and guardians, and information for parents on how to talk to their teens about intimate images. We also worked with Thorn and their NoFiltr brand to create and promote educational materials that reduce the shame and stigma surrounding intimate images, and empower teens to seek help and take back control if they have shared them or are experiencing sextortion.

We are also educating more people to avoid sharing child exploitation content, even in outrage or to raise awareness, as we know this type of sharing still causes harm. Research suggests that more than 75% of people that we reported to the National Center for Missing and Exploited Children (NCMEC) for sharing child exploitative content shared with no apparent intention of harm. Instead, users often share this content to raise the alarm or warn friends and family. Sharing this content violates our policies, regardless of intent and Facebook removes it as soon as it is detected. We also launched a global "Report it, Don't Share it" campaign reminding people of the harm caused by sharing this content and the importance of reporting this content. Finally, we have worked with public awareness experts to launch the Help Protect Children Campaign, which prevents the sharing of CSAM and encourages users to report such content instead. The campaign educates users on the harm CSAM causes, and the impact it has on victims. It encourages anyone who sees harmful videos and images of teens to protect the victim by reporting it immediately.

We also educate young people with in-app advice on avoiding unwanted interactions. We have seen tremendous success with our safety notices on Messenger, which are banners in our apps that provide tips on spotting suspicious activity and educating people on how to take action. These notices help people avoid scams, spot impersonations, and, most importantly, flag suspicious adults attempting to connect to minors.

Our work to safeguard young people extends to the broader internet. We recognize the power of cross-industry collaboration to create a child safety ecosystem. Since 2016, we have hosted regular child safety hackathons with NGOs. These events focus on coding and prototyping projects focused on making the internet a safer place for children. Likewise, our photo and video-matching technologies have been open source since 2019. By contributing back to the tech industry with this code, we hope to enable more companies to keep their services safe. In 2020, we collaborated with partners across the industry to establish Project Protect. This coalition is

designed to protect young people online and guide the work of the Technology Coalition for the next 15 years.

b. What are you proactively doing to educate the minors that use your services about online health and safety?

Please see the response to your Question 7(a).

Question 8. **Sextortion has become increasingly prevalent. Offenders may use grooming techniques or basic trickery to manipulate victims into providing nude or partially nude images of themselves, which are then used to coerce victims into sending more graphic images and videos or pay a ransom. These criminals often threaten to post the images or sensitive images publicly or send them to the victim's friends and family if the child does not comply. From May 2022 to October 2022, U.S. law enforcement and NCMEC witnessed an alarming increase in CyberTips and reports where minors have been sextorted for money. Many young boys, including in California, have committed suicide out of desperation, leaving their loved ones devastated.**

a. How is your company responding to the growing threat of financial sextortion?
b. What methods are in place to detect and disrupt this type of abuse in real time?
c. What kind of user education and awareness are you engaged in?
d. Are you aware of a higher prevalence of sexual extortion or abuse against certain demographics among young users? If not, will you commit to studying this issue and making that kind of information available to improve public education and protection measures?

Having a personal intimate image shared with others can be devastating, especially for young people. It can feel even worse when someone threatens to share that image if a person does not give more photos, sexual contact, or money—a crime in most jurisdictions, commonly referred to as sextortion.

At Meta, we take a multi-faceted approach to combat sextortion scams. These efforts include (i) strict policies against content or activity that sexually exploits or endangers children, including sextortion; (ii) human and machine detection and enforcement, including specialized teams focused on combating sextortion and automated rules that detect and action at scale accounts; (iii) proactive investigatory work, including targeted investigations and removal of violating accounts to disrupt networks of bad actors attempting to exploit or financially extort minors, and—when appropriate—reporting them to the National Center for Missing and Exploited Children (NCMEC); (iv) safeguards to help prevent suspicious adult accounts from finding or interacting with teens on our apps, including parental controls; and (v) provide education and awareness resources to those who may had their intimate images shared online. These efforts are described in more detail below.

We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. We work to prevent this content, as well as inappropriate interactions between young people and suspicious accounts attempting to take advantage of them. We also prohibit behavior that exploits people, including sharing or threatening to share someone's intimate images. In addition, impersonation is one way criminals gain the trust of their sextortion victims, which is one reason why our policies prohibit it. This helps address the problem at the root and prevent downstream harms, like sextortion. We have invested heavily in strengthening our technology to keep fake accounts off Facebook and Instagram and we cooperate with law enforcement and respond to lawful information requests in prosecutions of scammers.

We have specialized teams working on combating sextortion. These teams are constantly working to understand the unique combinations of on-platform behaviors used by criminals seeking to exploit our services. We build automation rules that allow us to detect and action—at scale and with high-precision—accounts committing financial sextortion. Our teams continue to work on new solutions to address sextortion industry-wide, including by developing new ways to identify people potentially engaging in sextortion and thwarting their efforts.

In addition, our dedicated teams investigate and remove these criminals and report them to authorities, including law enforcement and NCMEC, when appropriate. We work with partners, like NCMEC and the International Justice Mission, to help train law enforcement around the world to identify, investigate and respond to these types of cases. We have developed a streamlined online process through which we accept and review all legal requests from law enforcement. If we have reason to believe that a child is in immediate or imminent danger, we may proactively refer a case to local law enforcement (as well as report it to NCMEC) to help safeguard the child.

We also work to protect people from sextortion by preventing unwanted contact across our apps and in our messaging services, especially between adults and teens. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists, which can be used as a lever by people trying to sextort others. We also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We also restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram,

restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also introduced stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens can not be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We have developed ways to help people control their own experience. For example, people can choose who can message them, and can block anyone they do not want to hear from. People can report nude or sexual photos or videos of themselves or threats to share these images or videos to our apps or technologies to prevent them from being reshared. Our teams review reports 24/7 in more than 70 languages. We have articles in our Help Center that help people understand how to report this activity on Facebook,[42] Instagram,[43] and Messenger.[44]

Finally, we use technology to identify and prevent accounts exhibiting potentially suspicious behavior from finding, following, and interacting with teen accounts. Specifically, we work to ensure that teens are not recommended to potentially suspicious adult accounts, and potentially suspicious adult accounts are not recommended to anyone (including to teens or other potentially suspicious adult accounts). We review a combination of signals to find these accounts, such as if a teen blocks or reports an account, or if an account repeatedly searches for terms that may suggest suspicious behavior.[45] On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting suspicious activity and educating people about how to take action. These notices help people avoid scams, spot impersonations and, most importantly, flag potentially suspicious accounts attempting to connect to minors.

Because this is an industry-wide concern, we also direct people to various tools to use if people have nude or sexual photos or videos to help prevent them from being shared or reshared online. Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down,

[42] How do I report an abusive photo on Facebook? | Facebook Help Center
[43] How to Report Things | Instagram Help Center
[44] Reporting Conversations | Messenger Help Center (facebook.com)
[45] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Anyone seeking support and information related to sextortion can visit our education and awareness resources, including the Stop Sextortion resources, developed with Thorn. These resources include immediate actions parents and teens can take if users are experiencing sextortion, as well as expert tips for teens, parents and guardians, and information for parents on how to talk to their teens about intimate images. We also worked with Thorn and their NoFiltr brand to create and promote educational materials that reduce the shame and stigma surrounding intimate images, and empower teens to seek help and take back control if they have shared them or are experiencing sextortion. Additionally, since 2006, we have worked with suicide prevention experts to support the Meta community. For those who may post potential suicide and self-harm content, we use proactive detection technology to send this content to our teams for prioritized review. When someone searches for, or posts, content related to suicide, self-harm, eating disorders or body image issues, they will see a pop-up with tips and an easy way to connect to organizations like the National Alliance on Mental Illness (NAMI) in the US.

We also need Congress to pass legislation requiring operating-system level age verification requirements. That would allow services like Instagram to more quickly identify suspicious behavior, such as adults pretending to be minors, and remove them from the app entirely before they can even make contact with a teen—in addition to the work we have already been doing to prevent this contact. This also allows parents to oversee and approve their teen's online activity in one place. When a teen wants to download an app, app stores would be required to notify their parents. Where apps like ours offer age-appropriate features and settings, parents can help their teens use them. Until then, we require people to provide their age when signing up for accounts on our services, which helps us to provide teens with age-appropriate experiences.

For more information about our work combating sextortion and intimate image abuse, please see our dedicated page in Safety Center, linked here: https://about.meta.com/actions/safety/topics/bullying-harassment/ncii.

Question 9. **Young people need to be at the center of regulatory discussions, and they need to be at the table as products and services they use are designed.**

a. Are you engaging young adults and youth in your conversations and policies around Trust and Safety on the platform?

We want everyone who uses our services to have safe, positive, and age-appropriate experiences, and we approach all our work on child safety and teen mental health with this in mind. We build

comprehensive controls into our services, we work with parents, experts, and teens to get their input, and we engage with Congress about what else needs to be done.

We collect input from teens in multiple ways. For example, we have hosted programs with organizations such as Girls Scouts of America and the National Parent Teacher Association to create awareness of our safety tools and to get feedback from teens. We have also launched TTC Labs, a global co-design program, that invites young people, parents and experts to participate as collaborators in our design process, empowering them to have their say and ensuring our products meet their needs.

b. How do you proactively keep up to speed with the most pressing issues facing young people online?

We want parents to have the information to help their teens have a safe and positive experience on our services. We work closely with groups like ConnectSafely, ParentZone, and Net Family News to develop resources for parents and guardians to help them have meaningful and open conversations with their teens about being online. For example, our Family Center includes an education hub where parents and guardians can access resources from experts and review helpful articles, videos and tips on topics like how to talk to teens about social media. Parents can also watch video tutorials on how to use the new supervision tools available on Instagram today. And the Meta Quest Parent Education Hub includes a guide to our VR parental supervision tools from ConnectSafely to help parents discuss virtual reality with their teens. In the US, we have collaborated with The Child Mind Institute and ConnectSafely to publish a Parents Guide. It includes the latest safety tools and privacy settings, as well as a list of tips and conversation starters to help parents navigate discussions with their teens about their online presence. And in our Safety Center, we provide co-branded resources for parents from our collaboration with expert organizations.

As part of our work developing Messenger Kids, in addition to our research with thousands of parents, we engaged with over a dozen expert advisors in the areas of child development, online safety and children's media and technology who helped inform our approach. We have also had conversations around topics such as responsible online communication and parental controls with organizations like National PTA and Blue Star Families, where we heard firsthand how parents and caregivers approach raising children in today's digitally connected world.

We have also launched TTC Labs, a global co-design program, that invites parents, young people, and experts to participate as collaborators in our design process, empowering them to have their say and ensuring our products meet their needs. And through our partnership with Smart Design, we conducted co-design sessions with parents and teens, and consulted with

experts in the US, the UK, Ireland, Brazil, Japan and India, empowering them to provide input about how our services can meet their needs.

With respect to the research community, we have a global team responsible for helping to ensure that Meta remains a leader in online safety. We employ and work with researchers from backgrounds that include clinical psychology, child and developmental psychology, pediatrics research, public health, bioethics, education, anthropology, and communication. We also collaborate with top scholars to navigate various complex issues, including those related to well-being for people on Facebook and Instagram.

Additionally, Meta awards grants to external researchers to help us better understand how experiences on Facebook and Instagram relate to the safety and health of our community, including teen communities. We also publish and share papers with researchers on issues related to young people. For example, we have ongoing relationships with groups like the Aspen Institute and the Humanity Center, and we are a founding sponsor of the Digital Wellness Lab run jointly by Harvard University and Boston Children's Hospital. And because safety and well-being are not just Meta issues, but societal issues, we work with researchers in the field to look more broadly at youth experiences on mobile technology and social media, and how to better support youth as they transition through different stages of life.

We have a long track record of using research and close collaboration with our Safety Advisory Council, Youth Advisors, Suicide and Self-Injury Advisory Group, and additional experts and organizations to inform changes to our apps and provide resources for the people who use them. These relationships and our research efforts have been instrumental in helping develop a number of the tools and features described above, including Take a Break, Quiet Mode, Nudges, Hidden Words, and Restrict, among others.

Question 10. **For many children, an open dialogue about their internet habits is a best practice, and healthy. But not every child has a parent or a caregiver that is looking out for their best interest. For many kids who are abused, a caregiver or parent is their abuser. Additionally, for many young people, their parents' knowledge of their sexual orientation or their interest in exploring it, fundamentally puts them in jeopardy. Solving for these different needs across our young people at the scale of social media and internet applications is really vital.**

a. How have you designed your parental tools with this dynamic in mind?

Meta shares your concerns regarding the health, safety, and security of the people who use our platforms, including vulnerable communities. We want to help keep young people safe online, and supporting parents, guardians, and educators is one way we do this. Our work also includes providing baseline protections for all teens, as well as tools they can use themselves to help

manage their time, experiences and the content they see—irrespective of whether they use our parental supervision tools. Our mission is to give people the power to build community and bring the world closer together. But none of this is possible if people do not feel safe on our technologies.

In developing our parent supervision tools, we worked with teens, guardians, academics, policy makers, civil society, and subject matter experts to uncover insights and considerations. These partnerships helped inform our product decisions, including balancing parent and teen needs, developing transparency around these tools, and creating age-appropriate experiences. Recognizing young people's need for safe places to connect online, we also partner with LGBTQ+ safety and advocacy organizations around the world to help design policies and create tools that foster a safer online environment. This approach is always evolving, and input from the LGBTQ+ community online is critical to informing critical trends and helping us continually improve Meta's technologies and programs.

More broadly with respect to the LGBTQ+ community, one of the areas we have focused on is improving the transparency and supportiveness of our help center resources. For example, we have published dedicated Facebook and Instagram Help Center pages that focus on the subject of supporting authentic platform experiences and safety for the LGBTQ+ community. These Help Center pages address support for authentic representation on the platform. We have also worked in partnership with LGBTQ+ advocacy groups to develop a "Be Kind Online" guide, which includes bullying prevention and safety tips to tackle LGBTQ+ abuse online. This guide is available in the Resource section of the Facebook Safety Center. And we link to the Trevor Project as a way to get help in our Emotional Health hub—a centralized resource center on the Facebook app with tips and information from leading experts—and include them in our ongoing suicide prevention efforts. Relatedly, Meta's Civil Rights Team has developed a civil rights review process across Meta's technologies called Project Height to provide an analysis framework for product teams to assess potential civil rights concerns presented in new product launches.

***Question 11*. Meta published research on the intent behind the distribution of child sexual abuse material on your services which your company determined ranged from malicious to non-malicious. For example, some users would spread CSAM with the intent of calling attention to the abuse of a child. Regardless of intent, this material is illegal, it's distribution retraumatizes victims, and it is referred to NCMEC. In 2021, your company announced two tools you were testing to prevent distribution of this material. One was a pop-up for users that searched for terms on your apps associated with child exploitation. And the other was a safety alert that informs people who shared viral, meme child exploitative content about the harm it causes.**

a. What impact have these interventions had on preventing the distribution of CSAM on your services?

We are proud of the work our teams have done to improve online child safety, not just on our services but across the entire internet. We have around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone. We have built and shared tools for removing bad content across the internet, and we look at a wide range of signals to detect problematic behavior. We go beyond legal requirements and use sophisticated technology to proactively seek out abusive material, and as a result, we find and report more inappropriate content than anyone else in the industry.

We also continue to educate people to avoid sharing child exploitation content. As you reference in your question, we know that sometimes people repost sexual images and videos of children in outrage or to raise awareness, and understand that reposting such content, even without malicious intent, re-victimizes the child. In 2021, we launched a video campaign on Facebook called "Report It. Don't Share It." in partnership with child safety organizations to encourage people to stop and think before resharing those images online and to report them to us instead. We also show notices to people to not share these images or videos, directing them to reporting tools. Our teams have delivered several iterations of our campaign to educate people on the harm CSAM causes, which have been delivered both in-app and via our safety partner networks. This campaign currently lives in our Safety Center.

b. What lessons have you learned about prevention that could aid other technology companies?

We build technology specifically to help tackle some of the most serious online risks, and we share it to help our whole industry get better:

- We have developed technology that identifies accounts exhibiting potentially suspicious behavior, and we review a number of signals to find these adult accounts, such as if a teen blocks or reports an adult, or if someone repeatedly searches for terms that may suggest suspicious behavior.
- We built technology behind Lantern, the only program that allows participating companies to share data about people who break child safety rules.
- We were a founding member of Take It Down, the service that enables young people to prevent their nude images from being spread online. This is an important tool that a teen can use to protect against the threat of sextortion.
- In 2020, we joined Google, Microsoft, and fifteen other member companies of the Technology Coalition to launch Project Protect, a plan to combat online child sexual

abuse. Meta is an active member of the Technology Coalition, where members work together to drive critical advances in technology and adoption of best practices for keeping children safe online. In particular, the Technology Coalition focuses on information sharing, facilitating high-impact information, expertise, and knowledge sharing across industry to disrupt and help prevent online child sexual exploitation and abuse, including creating and expanding robust systems and processes to share information and threats about exploitative or predatory behaviors.

- We work closely with safety advisors and professionals, as well as leading online safety nonprofits and NGOs, to combat child sexual exploitation and aid its victims.

We have engaged with child-safety organizations and academic researchers to complete child-safety research that has helped move the industry forward. For example, we recently partnered with the Center for Open Science on a pilot program to share privacy-preserving social media data with academic researchers to study well-being.

Questions from Senator Tillis

Question 1. **Twenty-one is the minimum age to purchase highly regulated adult products such as alcohol, tobacco, and nicotine. Nevertheless, there is a proliferation of user-generated content posted on social media sites featuring underage use of these products.**

Recently, some have proposed banning these age-restricted products due in part to the user-generated content being available on your respective platforms. Surely, banning these products cannot be the answer. However, we must do more – your company must do more – to shield underage audiences from exposure to this content.

Therefore, as the content moderator of these platforms, what policies do you have in place, and what more can you do, to prevent this type of user-generated content from reaching underage audiences? How do you respond to requests to pull this content from your sites when deemed inappropriate for underage audiences?

We have strong and clear policies restricting the advertising or sale of alcohol and tobacco-related products, including e-cigarettes, on our services. We want teens to have safe, age-appropriate experiences on our apps. We have developed more than 50 tools and resources to support teens and their parents, and we have spent over a decade developing policies and technology to address content that breaks our rules or could be seen as sensitive. In January, we announced that we would automatically place all teens into our most restrictive content control setting and start to hide more types of content for teens on Instagram and Facebook. This includes content that does not violate our Restricted Goods and Services policy but which may come close.

Across both Facebook and Instagram, our policies distinguish between three types of content: organic content, including posts and images that people share and branded content; paid advertisements; and commerce listings, such as product listings in the Facebook Marketplace. Facebook's Community Standards and Instagram's Community Guidelines prohibit any organic content attempting to buy, sell, trade, donate, or gift alcohol or tobacco, with limited exceptions for brick-and-mortar and online retailers, detailed below. We have developed self-service tools to give companies the ability to age-gate their Facebook Pages and Instagram Business and Creator accounts, so that their organic content should only be shown to people above the age selected (for example, age-gating a Page to people 18 and older). Content that attempts to buy, sell or trade real life regulated goods, such as alcohol and tobacco, is also prohibited in Meta Horizon Worlds. And under our Recommendations Guidelines, content that promotes the use of certain regulated products, such as tobacco or vaping products, may not be eligible for recommendations on Instagram.

Brick-and-mortar and online retailers may promote restricted goods like tobacco and alcohol available for sale off of our services; however, we restrict visibility of this content for minors. We regularly consult with experts in adolescent development, psychology and mental health to help make our platforms safe and age-appropriate for young people, including improving our understanding of which types of content may be less appropriate for teens.

Question 2. **Public reports conclude that drug cartels use social media like TikTok, META, X, Snapchat, and others to plan, organize, and communicate in real-time. These communications coincide directly with criminal activity.**

What are your companies doing to crack down on cartel coordination? Specifically, in the recruitment of children to commit crimes or assist in the sale/distribution of illicit drugs?

Our policies prohibit criminal organizations from using Facebook and Instagram, and we remove these organizations from our platforms when we become aware of them. We will continue to take action against anyone, including cartels, who use our platforms in an attempt to organize the sale of illegal drugs.

Under our human exploitation policy, Meta has and will continue to forbid criminal organizations and other human smugglers from using our platforms to offer or facilitate their services. With respect to cartels, we work with law enforcement to obtain identifying information, and then we fan out systems to help us find instances of them across our platforms and take them down right away.

We expedite requests pertaining to child safety, and we have a team dedicated to engaging with NCMEC, International Centre for Missing & Exploited Children, Child Exploitation and Online Protection Command, Interpol, the FBI, and numerous other local, federal, and international law enforcement organizations and departments to ensure that they have the information and training needed to make the best use of this process and that we are supporting efforts to improve these processes. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC to help safeguard the child.

Question 3. **What steps does your platform take to proactively remove, delist, and ban any posts, users, websites, and advertisements associated with the sale and distribution of fentanyl and other illicit drugs?**

At Meta, we have in place multiple policies that prohibit drug-related content, including one related to high risk drug sales that we launched last year. Facebook's Community Standards and

Instagram's Community Guidelines prohibit buying, selling, or trading of high-risk drugs (defined as drugs that have a high potential for misuse, addiction, or are associated with serious health risks, including overdose; *e.g.*, cocaine, fentanyl, heroin), or non-medical (defined as drugs or substances that are not being used for an intended medical purpose or are used to achieve a high), or pharmaceutical drugs (defined as drugs that require a prescription or medical professionals to administer). And we have updated our Community Standards to make clear that our "non-medical drugs" prohibition includes precursor chemicals, like those that could potentially help manufacture dangerous drugs like fentanyl. We do not allow content that admits to buying, trading, or coordinating the trade of high-risk drugs or non-medical drugs personally or through others; or content that admits to personal use of high-risk drugs or non-medical drugs without acknowledgment of or reference to recovery, treatment, or other assistance to combat usage. We also prohibit content that speaks positively about, encourages use of, coordinates or provides instructions to make or use high-risk drugs or non-medical drugs. Our Advertising Standards prohibit ads that promote the sale or use of illicit or recreational drugs, or other unsafe substances, products, or supplements. And our Commerce Policy strictly prohibits listings that promote the buying or selling of drugs, drug paraphernalia, or prescription products.

We take a multi-pronged approach to enforce these policies, using sophisticated technology such as machine learning, reports from our community, and human review. Our technology helps us in two main areas: (i) proactive detection and (ii) automation. Artificial intelligence has improved to the point that it can proactively detect violations across a wide variety of areas, including drug-related content, without relying solely on community reports and often with greater accuracy. With automated enforcement, we sometimes require human review to understand the context in which a piece of content was posted (for example, to ensure it was not posted in the context of education or awareness-raising). We use machine learning to scale the work of our content reviewers. By allowing our AI systems to action content that is highly likely to be violating, this helps scale content decisions without sacrificing accuracy so that our reviewers can focus on decisions where more expertise is needed to understand the context and nuances of a particular situation. In the fourth quarter of 2023, of the drug-sales violating content we removed, over 97% was detected before a user reported it on Facebook and over 99% on Instagram.

With respect to discoverability, we also block and filter hundreds of terms associated with illicit drug sales, and we continue to review additional hashtags to detect violations of our policies. When people search Facebook and Instagram for drug-related hashtags and search terms we have identified—including information on opioids—we surface a pop-up interstitial that directs them to resources on the Substance Abuse and Mental Health Services Administration (SAMHSA) National Helpline and other resources for free and confidential treatment and education.
We hold people increasingly accountable for violating content they post on Facebook and Instagram. Under our high-risk drugs policy, one violation will result in the disabling of an

account. For most other violations, we count repeat violations and will disable accounts that repeatedly violate our policies. This means that if we are made aware that a person continues to post content that goes against the Facebook Community Standards or Instagram Community Guidelines, despite warnings and restrictions, we will disable the person's account. Additionally, beyond our strike policy, we also disable some accounts when we become aware of them, such as those of dangerous individuals and accounts created to get around our restrictions. More specifically, with respect to restricted goods such as illicit drugs, we also remove accounts on Facebook and Instagram we determine are dedicated to the sale of such goods.

Last year, we took a number of steps to enhance our detection and enforcement of this content. We have developed new detection pipelines to not only identify violating posts on our platform, but also to disable individuals responsible for posting violating content. This work includes identifying sales of known fentanyl precursors that we received from the International Narcotics Control Board, and building detection measures based on a variety of technical signals. We are also working to detect and remove very large networks of "non-delivery" scammers, who are masquerading as drug dealers in order to try to defraud people, but who never deliver any illicit substances.

In addition to our efforts to keep this content off our platform, we are committed to working with law enforcement to support the work they do to keep us safe. When law enforcement alerts us about illegal drug-related activity on Facebook or Instagram, we work to mitigate that threat. We have developed tools designed to quickly respond to law enforcement requests submitted in connection with official criminal investigations. We have a dedicated, trained Law Enforcement Response Team that reviews and evaluates government requests for user data individually, whether the request is related to an emergency or through a legal process initiated by law enforcement. We also contact law enforcement proactively if we become aware of a credible threat of harm. We handle disclosures to law enforcement on a case-by-case basis, and such disclosures include threats related to drug trafficking and fentanyl-laced counterfeit pills. Finally, we are working with DEA to better understand evolving tactics and emerging threats in this space.

The unprecedented public health crisis relating to non-medical synthetic drugs—especially fentanyl—has impacted so many, often with tragic results. This is why it requires a whole-of-society approach, working together to strengthen our ability to respond to this crisis. We have and will continue to collaborate with others—including government, and specifically law enforcement, health experts, researchers, our peers at other tech companies, and grassroots recovery and support organizations—to tackle these issues. We lead efforts, alongside Snapchat, to enhance the effectiveness of drug-related signal sharing among industry partners and work to recruit additional members to the Anti-Illicit Drugs Signal Sharing program, through which participating organizations can share data to help mitigate the threat. We also built the

technology upon which this program runs, an API called ThreatExchange. This work strengthens our ability to find and remove illicit drugs if they come onto our platforms. We will look for ways to continue to plan to extend our information sharing, including additional signals such as emojis that have proved successful for Meta in detecting high-risk drug sales. As the program continues, we hope additional companies are willing to partner with us to combat this industry-wide issue. We also have a long history in the US of developing programs and partnerships that help raise awareness about the national overdose and fentanyl crisis, promote education, and connect individuals and families with resources and help. We are committed to working with local communities, national organizations and government leaders to fight this epidemic.

Question 4. **One area of growing concern is the sale and distribution of fake or counterfeit vaping devices online, particularly in connection with so-called Delta-8 THC. Counterfeit vapes, many coming from China, have unsafe and even potentially deadly chemicals. They have caused hospitalizations and death. What are your platforms doing to combat this problem?**

For more information on restricted content, please see the response to your Question 1. For more information on detection and enforcement of our content moderation policies, please see the response to your Question 3.

In addition, our Advertising Policies prohibit certain advertisements for any person, regardless of age, including ads that promote the sale or use of tobacco products and related paraphernalia or ads that promote electronic cigarettes, vaporizers, or any other products that simulate smoking. E-cigarettes have always been covered by this policy, but to enhance its clarity, we updated the policy in December 2019 to explicitly prohibit ads for e-cigarettes and vaping. Our Branded Content Policies, which apply to organic content posted by an influencer working with a company to promote their product, also prohibit the promotion of tobacco-related products, including e-cigarettes. In commerce listings, we also prohibit the sale of tobacco-related products, including e-cigarettes.

Question 5. **What are the main impediments your platform encounters in identifying all fentanyl and illicit drug advertisements posted to your platform(s) automatically? Please describe any circumstances in which you do not or cannot apply detection technologies against content transmitted on your platform(s).**

For more information on detection and enforcement of our content moderation policies, please see the response to your Question 3.

Question 6. **How many posts, users, websites, and advertisements have you removed, delisted, and banned per year for the sale and distribution of fentanyl and other illicit drugs? How many per year? Have you seen an increase in illicit drugs being advertised to children on your platform(s)?**

Views of violating content that contain restricted goods and services, like illicit drugs, are typically infrequent, as we remove much of this content before people see it. We publicly report the amount of content we action on Facebook and Instagram for violating our policies on a quarterly basis in our Community Standards Enforcement Report, available at https://transparency.facebook.com/community-standards-enforcement. From January to December 2023, we removed approximately 9.3 million pieces of content related to drugs on Facebook and about 10.1 million pieces of content related to drugs on Instagram.

For more information on detection and enforcement of our content moderation policies, as well as our advertising standards, please see the response to your Question 3.

Question 7. **Are there any other roadblocks or impediments that you face in addressing fentanyl and illicit drug advertisements on your platform(s), and working with law enforcement on such matters? If yes, what are they? If no, how many cases have been transmitted to law enforcement and DEA?**

Meta responds to government requests for data in accordance with applicable law and our terms of service. All requests we receive are carefully reviewed for legal sufficiency and we may reject or require greater specificity on requests that appear overly broad or vague. Data on the number of requests we received, the number of accounts requested, and the rate we complied with all or some of the government's requests going back to 2013 is available in our Transparency Center.

For more information, please see the response to your Question 3.

Question 8. **How do you work with organizations, advocates, and experts focused on drug prevention and addiction recovery to adapt your products and operations to keep up with the illicit drug crisis — including working with parents that have lost children due to lethal drugs bought online?**

We are aware of the acute need to focus on this issue given the unprecedented public health crisis around non-medical synthetic drugs, especially fentanyl. And we understand and share your concern about the public safety and health threat it poses. We know this problem impacts so many, often with tragic results, which is why it requires a whole-of-society approach. When we work together, it strengthens our ability to respond to this crisis.

That is why we collaborate with others—including government, health experts, researchers, our peers at other tech companies, and grassroots recovery and support organizations—to tackle these issues. For example, since 2022, we have been in an information-sharing program with Snapchat that helps both platforms identify patterns and signs of illicit drug-related content and activity. We will look for ways to continue to plan to extend our information-sharing, including additional signals such as emojis that have proved successful for Meta in detecting high-risk drug sales. As the program continues, we hope additional companies are willing to engage and partner to help protect people and combat this industry-wide issue.

In July 2023, the State Department launched the Global Coalition to Address Synthetic Drug Threats aimed at uniting countries worldwide in a concerted effort to prevent the illicit manufacture and trafficking of synthetic drugs, identify emerging drug trends, and respond effectively to their public health impacts. We are also working with the State Department, the UN, and Snap to explore the feasibility of standing up a Tech Cooperation on Synthetic Drugs initiative (similar to Tech Against Terror and the Global Internet Forum to Counter Terrorism). The purpose of this effort is to cooperate and share best practices among industry, civil society, and governments on drug traffickers' use of the Internet and to develop awareness programs and campaigns against fraudulent, dangerous drugs sold online.

We care deeply about the impact of drug addiction in our communities, and are committed to continuing to do our part to combat this epidemic. Meta has a long history in the US of working with leading experts and non-profit organizations on programs that aim to address the national overdose and fentanyl crisis by raising awareness, promoting education, and connecting individuals and families with resources and help. Examples of these partnerships include:

- Song For Charlie (SFC), a leading non-profit working to raise awareness of the fentapill (i.e., fake pills made of fentanyl) crisis. We partnered again with SFC this year to support the second annual Senate-designated and DEA-recognized Fentanyl Awareness Day, helping SFC expand their reach.
- The Ad Council is continuing Drop the F-Bomb, a parent-focused campaign emphasizing the prevalence and dangers of fentanyl on Fentanyl Awareness Day. This campaign mobilizes and equips parents and caregivers to begin candid discussions with their families about the drug. We led in the creative development of this campaign, which provided parents with resources like Fentanyl 101 facts and guides on how parents can educate their families on the dangers of fentanyl. According to the Ad Council, the campaign reached nearly 8 million parents on our platforms in 2023.
- Mobilize Recovery, an organization that brings local leaders together to organize community engagement for people in recovery, family members, and recovery allies, hosted a series of regional events leading up to a Meta co-hosted national conference in Washington, DC in September 2023. The regional events offered an opportunity to listen

to community leaders who are on the front lines of our national overdose crisis, providing recovery support services, prevention education in schools and transitional housing to those in early recovery.

- We partnered with the Center for Safe Internet Pharmacies (CSIP) for the sixth straight time to support DEA Prescription Drug Takeback Day by connecting people with drop-off locations.
- We have partnered with Partnership to End Addiction, a leading nonprofit working to transform how the nation addresses addiction, on campaigns to help connect parents, guardians and young people with educational resources on prevention and recovery. According to the Partnership to End Addiction, in H2 2023 alone, our campaign reached more than 10 million people with recovery resources in both English and Spanish across our platforms and drove 35,000 people to Partnership to End Addiction's English and Spanish Risk Assessment tools, which help family members identify risk factors specific to their loved ones and provide personalized guidance on how to mitigate and address these risks.
- We also worked with the Partnership to End Addiction to launch the Stop Opioid Silence campaign, a national public awareness campaign aimed at breaking down the stigma and shame associated with opioid use disorders. We partnered with over 150 members of Congress to add their voices to the campaign with public service announcement videos that reached more than 70 million people on our services.

Thanks to expert feedback, we know how vital it is to give people—especially anyone personally impacted by this issue—platforms where they can feel safe to discuss the dangers of drugs and the ways to overcome addiction. That is why our policies allow people to talk about their recovery or that of a loved one to raise awareness, provide education, and connect to resources that can help.

***Question 9*. What are the total number of meetings that your company has had with parents to address online safety concerns? Can you provide the total number of meetings over the last three years? Please separate this last question's answer by number per year.**

We reach parents in a variety of ways, including through Family Center's Education Hub, advertising campaigns, in-app promotion, various roundtables and other events with parent groups, and our ongoing work with safety partners. As a few notable examples, we work closely with external groups such as ConnectSafely and Net Family News to develop resources for parents and guardians to help them have meaningful and open conversations with their teens about being online. And through our partnership with Smart Design, we conducted co-design sessions with teens and parents and consulted with experts in the US, the UK, Ireland, Brazil, Japan and India. That co-design work invites young people, parents and experts to participate as

collaborators in our design process, empowering them to provide input about how our services can meet their needs. Another example of these efforts is our launching of TTC Labs, a global co-design program, that invites young people, parents and experts to participate as collaborators in our design process, empowering them to have their say and ensuring our products meet their needs.

We have also hosted community events with parents to educate them on the tools and resources we offer, including six events in 2023, in Seattle, Chicago, Los Angeles, Nashville, Miami, and New York. Additionally, in 2023, we hosted "Screen Smart" events in six cities (NYC, LA, Miami, Chicago, Nashville, and Seattle), bringing together over 250+ parenting influencers and local stakeholders to educate them on the tools, features, and resources we provide to support parents and teens. Attendees shared content and information from the events about our tools, features, and resources, and those pieces of content reached nearly 50 million impressions. And we recently announced a series of Screen Smart workshops to help empower parents to confidently manage their teens' usage of smartphones and devices—including on Meta's platforms. Our ads encouraging parents to use our youth well-being tools and features were seen more than one billion times by tens of million people in the United States since August 2022.

***Question 10*. In 2022, then National Center for Missing & Exploited Children (NCMEC) received over 32 million reports of Child Sexual Abuse Material (CSAM). Reports of online sex crimes to the CyberTipline are growing exponentially year by year. Out of those 32 million reports, how many did your platform submit to NCMEC?**

In 2022, we made over 26 million reports between Facebook, Instagram, and WhatsApp. The rest of the industry made less than 5 million reports collectively. We expect to continue providing more reports to law enforcement than our peers, thanks to our industry-leading work on keeping people safe. For example, WhatsApp—which has long been encrypted—removes hundreds of thousands of accounts per month for CSAM violations. In 2022, WhatsApp also made over one million reports to NCMEC, all without breaking encryption. This was significantly more than all other encrypted messaging services combined. NCMEC has acknowledged that Meta continues to be an industry leader in this work and that Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs." We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

***Question 11*. There is concern that this number is going to fall dramatically this year because of the adoption of end-to-end encryption, not because the problem is going away. How will your company track and address this issue moving forward?**

Child exploitation is a horrific crime that we work aggressively to fight on and off our platforms. We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. We have processes in place to remove policy-violating content, regardless of the context or the person's motivation for sharing it. We also have developed aggressive, cutting-edge technology to prevent, find, remove, and report policy violating content. In addition to this technology, we have invested in specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to both find and review potentially violating content, accounts and adversarial networks.

As a general matter, we do not share detailed descriptions of how our tools work or our enforcement efforts, which, if revealed, could provide a roadmap to highly-motivated bad actors who seek to evade our detection and NCMEC reports, which would ultimately undermine our efforts. That said, we are working hard to further augment the measures we have in place to strive to evolve as predatory behaviors and coded language do as well.

For example, we work to promptly disable accounts on Facebook and Instagram for various violations of our child exploitation policies, such as the apparent malicious[46] distribution of CSAM or sexual solicitation of children. We also recognize that predators may attempt to set up multiple accounts to evade enforcement of our policies. That is why when we disable accounts for these severe violations, we also work to disable explicitly linked accounts (where a person has linked their Facebook and Instagram profiles), high-confidence linked accounts (where we have high confidence that the same person is using multiple accounts), and restrict those devices from setting up future accounts. We also utilize technology and teams to detect and eliminate abusive networks to take on predators who attempt to use our services to connect online. We also collaborate with industry on new programs, such as with Take It Down and Lantern, of which we are founding members. Lantern is a program that enables technology companies to share signals of a child safety threat and Take It Down is a platform designed to proactively prevent young people's intimate images from spreading online. Importantly, we want teens to have safe, age-appropriate experiences on our apps, including on our encrypted services. We do not believe moving to an encrypted messaging environment means sacrificing safety. It is already widely used by other large messaging services to protect people's private messages and provide people with the privacy and security they expect when messaging friends and family. That is why we will continue to support encryption, while putting features in place to help keep people safe.

[46] We distinguish between "malicious" and "nonmalicious." In the "malicious" group are people we believe intended to harm children with their content, and in the "nonmalicious" group are people we believe, based on contextual clues and other behaviors, likely did not intend to cause harm to children. (For example, they shared the content with an expression of abhorrence.) Regardless, and consistent with US law, these actors are reported to NCMEC.

With respect to Messenger and Instagram Direct Messages, as we rollout end-to-end encryption our approach to safe encrypted experienced is focused on three key elements: (i) preventing potential harm in the first place; (ii) giving people ways to control their experience; and (iii) responding to violations of our policies quickly. This approach is detailed in our whitepaper, Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging.

To address the potential for harm, we have built tools and policies specifically to help young people manage interactions with adults. For example, on Facebook and Instagram, we do not recommend to anyone, through Facebook's "People You May Know" algorithm or otherwise, accounts we identify as exhibiting potentially suspicious behavior.[47] Specifically, teens are not recommended to adult accounts exhibiting potentially suspicious behavior, and adult accounts exhibiting potentially suspicious behavior are not recommended to anyone (including teens and other potentially suspicious adult accounts). Furthermore, accounts for people under 16 (or under 18 in certain countries) are defaulted to private, so teens can control who sees or responds to their content.

On Facebook and Instagram, we recently announced additional steps to help protect teens from unwanted contact, turning off their ability to receive messages from anyone they do not follow or are not connected to on Instagram—including other teens—by default. We restrict adults over 19 from sending private messages to teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos or repeated messages from people they do not know. Similarly, on Messenger, teens will soon only receive messages from Facebook friends or people they are connected to through phone contacts by default.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

[47] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

In addition, in an end-to-end encrypted environment, we use machine learning to proactively detect accounts engaged in malicious patterns of behavior. Our machine learning technology will look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors.

We have also built more than 50 tools, resources, and features to help protect teens. For more information on these tools as well as to review resources from experts, visit our Family Center to learn more: https://familycenter.meta.com/.

To help us respond to violations of our policies quickly, we encourage people to report messages to us in both encrypted and unencrypted services. We have made our reporting tools easier to find and started encouraging teens to report at relevant moments, such as when they block someone. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting potentially suspicious activity and educating people on how to take action. These notices help people avoid scams, spot impersonations and flag accounts that have been exhibiting potentially suspicious behavior that attempt to connect to minors.

Keeping young people safe online has been a challenge since the advent of the internet. Online predators are determined criminals who use multiple apps and websites to target young people. They also test each platform's defenses, and they learn to quickly adapt. That is why now, as much as ever, we are working hard to stay ahead of these threats by developing technology to root out predators, work with specialists dedicated to online child safety, and share information with our industry peers and law enforcement.

***Question 12*. Has your platform seen an increase of suspected online child sexual exploitation-CSAM over the past few years? If so, what do you believe is the driving factor on why it's happening on your platform?**

We publish the Community Standard Enforcement Report to more effectively track our progress and demonstrate our continued commitment to making Facebook and Instagram safe and inclusive. Updated metrics and recent trends organized by policy area and extending back over five years are available at the Meta Transparency Center.

There are many factors that contribute to increases or decreases in the amount of violative content reported on platforms, including improvements in detection. We track prevalence, which tells us how often content that violates our standards is seen relative to the total amount of times any content is seen on Facebook or Instagram. We estimate the prevalence of this content on Instagram to be below 0.01%.

***Question 13*. What are some new tools or strategies that your platform has implemented to identify CSAM? How closely does your platform work with NCMEC?**

We have invested heavily in sophisticated technology that helps us proactively find and remove violating content and accounts. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in enqueuing certain content for human review. For example, we use technology designed to proactively find child exploitative imagery to identify and prioritize reports of content that are more likely to contain content that violates our child safety policies. For example, we use AI and machine learning to proactively detect and take action against known child exploitation and sexualizing content, leveraging technology available across industry for CSAM hash matching, including methods that Meta developed and open sourced. We also detect novel, previously unknown child exploitation and sexualizing content using proprietary detection technology, in conjunction with a team of specialized human reviewers. More specifically, we use PhotoDNA and other photo- and video-matching technologies that detect identical or near-identical photos and videos of known child exploitative content, and we use Google's Content Safety API to help us better prioritize content that may contain child exploitation for our content reviewers to assess. We also use technology to detect and remove Instagram Reels and Stories that violate our Community Guidelines, including by scanning for CSAM and for CSE indicators.

Beyond technology, we also continue to encourage user reporting, as it can provide context that may be helpful in taking action against potential violations of our policies, including as related to CSAM. On Facebook and Instagram, we enable people to report content or conduct they believe violates our policies and flag for our review. We have built systems and review processes to prioritize and appropriately action violating content or accounts and, when appropriate, report it to NCMEC or law enforcement.

Our work in this space is not limited to our own services. We are proud of the strong relationship we have developed with NCMEC and continue to report all apparent CSAM found globally to NCMEC's CyberTipline across our family of apps. As NCMEC noted recently, Meta goes "above and beyond to make sure that there are no portions of their network where this type of activity occurs." We also are deeply committed to improving the entire ecosystem and have engaged with child safety nonprofits and academic researchers to complete child safety research with fieldwide impact. Additionally, to help safety stakeholders identify and respond to the most high priority reports at NCMEC, we funded and helped rebuild their case management tool to ensure investigators can get to the most important cases quickly.

In addition, Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

Question 14. **What resources or help does your platform provide to victims of CSAM? Does your platform work with local victim groups and professionals?**

We consult with a number of external experts and partners—including survivors and survivor organizations—as we work to provide people with a safe and positive experience on our services. As further described below, this includes other members of the technology industry, nonprofits, law enforcement, civil society organizations, and academics with relevant experience.

More specifically, we incorporate feedback from survivors in a number of ways, including collaborating with organizations who work with survivors in a safe, trauma-informed way, and meeting survivors at conferences hosted by various stakeholders. Meta also provides funding to NCMEC's free service to help survivors and families impacted by online sexual exploitation, and maintains a dedicated reporting channel where NCMEC staff are able to flag non-CSAM posts and profiles that threaten or otherwise identify CSAM survivors for our review and action.

In addition, to combat child exploitation both on and off our platforms, we work regularly with child safety professionals to help us understand evolutions in coded language and to identify new and evolving terms, phrases, slang, and emojis that could be used in an attempt to evade our detection systems and bypass our policies.

We also work with these professionals and organizations to build various interventions, including but not limited to our search interventions, safety notices, and safety education campaigns. We have also worked with child safety researchers to conduct collaborative research to improve child safety protections on our platforms.

Our collaborative work to address child safety does not stop with improving our own services. We also are deeply committed to improving the entire ecosystem and have engaged with child safety nonprofits and academic researchers to complete child safety research with fieldwide impact. Our efforts with these professionals also include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, among other things. Additionally, to help safety stakeholders identify and respond to the most high priority reports at NCMEC, we funded and helped rebuild their case management tool to ensure investigators can get to the most important cases quickly.

In addition, Instagram and Facebook are founding members of Take It Down—a service by NCMEC designed to proactively prevent young people's intimate images, including AI-generated content, from spreading online. We provided financial support to NCMEC to develop Take It Down, building on the success of StopNCII.org, a platform we developed that helps adults stop the spread of their intimate images online. We have made both Take It Down and StopNCII easily accessible on our apps when people are reporting potentially violating content.

We have also worked with Thorn, a nonprofit that builds technology to defend children from sexual abuse, to develop updated guidance for teens on how to take back control if someone is sextorting them. It also includes advice for parents and teachers on how to support their teens or students if they are affected by these scams. These resources can be found in our updated Sextortion hub within Meta's Safety Center.

***Question 15*. What are the top technical hurdles your company faces in combatting CSAM?**

Keeping young people safe online has been an industry-wide challenge since the start of the internet. Online predators are determined criminals who use multiple apps and websites to target young people and find each other. They also test each platform's defenses, and they learn to quickly adapt. Some of the key issues throughout the industry that arise due to the rapidly evolving landscape include the ability to train our technology to better detect harmful content

like CSAM, as well as identifying coded—and often ordinarily benign—language that adversarial actors may use to hide their harmful activity.

As threats from criminals evolve, we have to evolve our defenses. That is why, in addition to developing technology that roots out predators, we hire specialists dedicated to online child safety and we share information with our industry peers and law enforcement. Still, no matter how much we invest or how effective our tools are, this is an adversarial space. There is always more to learn and more improvements to make.

It can also be challenging for parents to supervise the many apps that their teens may use, which is one of the reasons we support federal legislation at the app store level that would make it simpler for parents to oversee their teens' online lives. Parents want to be involved in their teen's online lives, and recent Pew research suggests that 81% of US adults support requiring parental consent for teens to create a social media account. But technology is constantly changing and keeping up with all the apps teens use can feel impossible. As an industry, we should come together with lawmakers to create simple, efficient ways for parents to oversee their teens' online experiences.

That is why we also support federal legislation that would make it simpler for parents to oversee their teens' online lives, including legislation that would require app stores to get parents' approval when teens under 16 download an app. According to a recent Morning Consult poll,[48] parents across both sides of the aisle overwhelmingly support this approach. 81% of Democratic-leaning and 79% of Republican-leaning parents back federal legislation for parental approval of teen app downloads. Over 75% of parents prefer app stores as more secure and straightforward venues for approving downloads, and a more effective method than individual app-level.

In addition to offering a simpler way for parents to approve their teens' app downloads, federal legislation needs to create standards for all apps to adhere to in areas like age-appropriate content, age verification, and parental controls.

We want to help find workable solutions and earlier this year we proposed a framework for legislation.[49] We designed this framework to create clear, consistent standards for all apps, to empower parents and guardians, and to preserve user privacy, in ways that are technologically feasible for the industry. This framework would:

- **Require app stores to get parental approval for teens under 16 to download an app.** Empowering parents to approve their teens' app downloads ensures that they oversee their teens' online experience. Placing the point of approval within the app store

[48] Morning Consult Survey
[49] A framework for legislation to support parents and protect teens online (January 16, 2024)

simplifies the process and leverages optional approval systems already offered by app stores. App stores would notify parents and request their approval when their teen wants to download an app, including Instagram.

- **Require certain apps, including social media apps, to offer supervision tools for teens under 16 that parents can activate and control.** Parents should have the tools they need to guide and support their teens online. Certain apps, including social media apps, should be required to offer some form of parental supervision tools, including the ability to set daily time limits on teens' usage, see which accounts their teen is following or friends with, and more. Furthermore, apps can quickly and easily implement these tools if a parent relationship is established in the app store.

- **Require app stores to verify age and provide apps and developers with this information.** Knowing a person's age helps ensure that apps can easily place teens in the right experience for their age group, but parents and teens should not have to provide sensitive information like government IDs to hundreds of apps to verify their age. Parents already provide this information when they purchase a teen's phone and set up their teen's account. App stores have this information and not only can they ease the burden on parents by sharing it with apps, they can help ensure teens are placed in age-appropriate experiences.

- **Require industry to develop consistent age-appropriate content standards across the apps teens use.** Parents are eager to have a better understanding of the content available to their teens and to have guidelines to help them evaluate whether an app is appropriate for their child. We need broader alignment across industry on the types of content companies should consider age appropriate, as there is for other media like movies and video games. It is time we have common industry standards for what is age-appropriate that parents can rely on.

- **Establish national standards to unify the complicated patchwork of inconsistent state laws, and that apply to all apps consistently.** Parents expect consistent standards across all the apps their teens access—regardless of where their teens access or use them.

- **Require industry to develop ads targeting and delivery standards that, for example, limit the personalization of ads for teens under 16 to age and location only.** Industry standards on ad targeting and delivery can help to ensure teens see relevant ads for age-appropriate products and services in their community (e.g., a college prep course) while eliminating the ability to target this audience based on online behaviors or activity. Personalizing ads by age and location is common across industries: for example, advertisers may place relevant ads during teens' TV shows, or in magazines or newspaper sections designed for teens.

Question 16. **There seem to be competing views on how to regulate algorithms. Some suggest that more transparency is needed, while others want more privacy. Can you provide your perspective on whether more or less transparency is needed when it comes to algorithms?**

We believe a better understanding of the relationship between people and the algorithms is in everyone's interest. With rapid advances taking place with powerful technologies like generative AI, it is understandable that people are both excited by the possibilities and concerned about the risks. Generally speaking, we believe that as these technologies are developed, companies should be more open about how their systems work and collaborate openly across industry, government, and civil society to help ensure they are developed responsibly.

Meta has taken concrete steps to enhance transparency regarding its algorithms. An approach to transparency Meta has been developing and advocating for some time is the publication of system cards, which give people insight into how our systems work in a way that is accessible for those who do not have deep technical knowledge. We have released a number of system cards for Facebook and Instagram to date. They give information about how our AI systems rank content, some of the predictions each system makes to determine what content might be most relevant to you, as well as the controls you can use to help customize your experience. They cover Feed, Stories, Reels, and other surfaces where people go to find content from the accounts or people they follow. The system cards also cover AI systems that recommend "unconnected" content from people, Groups, or accounts they do not follow.

We have also shared the types of inputs—known as signals—as well as descriptions of the predictive models these signals inform that help determine what content you will find most relevant from your network on Facebook. The categories of signals we have released represent the vast majority of signals currently used in Facebook Feed ranking for this content. You can find these signals and predictions in the Transparency Center, along with how frequently they tend to be used in the overall ranking process.

We also have made it possible to see details directly in our apps about why our systems predicted content would be relevant to you, and the types of activity and inputs that may have led to that prediction. We have expanded our "Why Am I Seeing This?" feature in Instagram Reels tab and Explore, and Facebook Reels, after previously launching it for some Feed content and all ads on both Facebook and Instagram. People are able to click on an individual reel to see more information about how their previous activity may have informed the machine learning models that shape and deliver the Reels they see.

Question 17. **Do you believe that large companies and platforms like yours can use algorithms to stifle innovation or small businesses?**

As the internet has grown over the last 25 years, the ways in which people share and communicate have exploded thanks to dynamic competition. The most successful platforms mature and adapt to people's changing preferences. Our services became and remain popular for this very reason—we constantly evolve, innovate and invest in better experiences for people against world-class competitors. We innovate and improve constantly because we have to.

Recent breakthroughs in AI, and generative AI in particular, have captured the public's imagination and demonstrated what those developing these technologies have long known—they have the potential to help people do incredible things, create a new era of economic and social opportunities, and give individuals, creators, and businesses new ways to express themselves and connect with people.

Indeed, our platforms are good for small businesses. Personalized ads, for example, play an important role for small businesses. Small businesses may not be able to afford the broad, mass marketing campaigns that big brands can. For many small and medium-sized businesses, personalized advertising is the secret ingredient that makes their success possible.

We face fierce competition for every service we offer. We compete hard, and we compete fairly and we are confident that our work in this space will enhance competition.

Question 18. **What do you believe is the role of government in regulating algorithms? What, if any, unintended consequences would there be if Congress gets involved?**

One way Meta helps people to build community is by using algorithms to recommend connections and content people might be interested in—for example, new Facebook Groups they might want to join, Pages they might like, or events they might want to attend—and by ranking content so that they are more likely to see the posts they care most about. This technology also helps protect our community by filtering, blocking, and reducing the spread of content that violates our policies or is otherwise problematic. We also use algorithms to help teens have age-appropriate experiences on our apps. This includes using them to filter out content that might be sensitive, and regularly recommending that teens update their privacy settings.

Generally speaking, we believe that as these technologies are developed, companies should be more open about how their systems work and collaborate openly across industry, government and civil society to help ensure they are developed responsibly. That includes giving people more insight into, and control over, the content they see. One model of transparency Meta has been developing and advocating for some time is the publication of system cards, which give people insight into how systems work in a way that is accessible for those who do not have deep technical knowledge. We have released a number of system cards for Facebook and Instagram to date, which give information about how our AI systems rank content, some of the predictions

each system makes to determine what content might be most relevant to people using our services, as well as the controls people can use to help customize their experience. The system cards cover Feed, Stories, Reels and other surfaces where people go to find content from the accounts or people they follow. The system cards also cover AI systems that recommend "unconnected" content from people, groups, or accounts they do not follow.

Congress could also bring more transparency, accountability, and oversight to the processes by which large internet companies make and enforce rules about what users can do or say on their services. We support efforts to bring greater transparency to algorithmic systems, offer people more control over their experience and require audits of services' content moderation systems—which, of course, include algorithms. We also support standards-setting processes that tackle questions like how to measure "bias" in an algorithm that—once established—could be required across the industry.

People of all political persuasions want large companies to take responsibility for combating illegal content and activity on their services. And when they remove harmful content, people want them to do so fairly and transparently. The sheer volume of user-generated content on the internet means that online companies have to make decisions about how to organize, prioritize, and deprioritize this content in ways that are useful to people and advertisers, while enforcing our policies against harmful content. Online services should be granted continued protection from liability for the content they carry if they can demonstrate that they have robust practices for identifying illegal content and quickly removing it. While it would be impractical to hold companies liable if a particular piece of content evades detection, they should be required to follow common industry standards and best practices.

Question 19. **Are you aware of your platform using surveillance advertisements to target children (anyone under the age of 18) with specific ads? If so, in your opinion, how can this be mitigated?**

We recognize that teens are not necessarily as equipped as adults to make decisions about how their online data is used for advertising, particularly when it comes to showing them products available to purchase. For that reason, we restrict the options advertisers have to reach teens, as well as the information we use to show ads to teens. We prohibit children under the age of 13 on any of our services that run advertising.

Last year, we made changes to how advertisers can reach teens, which included removing the ability for advertisers to target teens based on their gender, interest, and activities. Accordingly, currently age and geography are the only information about a teen that we use to show them ads. In addition to these restrictions, we also provide teen-specific controls to help them understand how ads work and the reasons why they see certain ads on our apps. Teens are able to manage

the types of ads they see on Facebook and Instagram with Ad Topic Controls. As noted, our Advertising Standards already prohibit ads about restricted topics—like alcohol, financial products and weight loss products and services—to be shown to people under 18 (and older in certain countries). But even when an ad complies with our policies, teens may want to see fewer ads like it. For example, if a teen wants to see fewer ads about a genre of TV show or an upcoming sports season, they should be able to tell us that. Teens can continue to choose to hide any or all ads from a specific advertiser. The topics we already restrict in our policies will be defaulted to See Less, so that teens cannot choose to opt into content that may not be age-appropriate.

In addition to these controls, we have also introduced more teen-specific resources to help teens understand how ads work and the reasons why they see certain ads on our apps. These changes reflect research and direct feedback from parents and child developmental experts. For example, we added a privacy page in 2023 with more information for teens about the tools and privacy settings they can use across our technologies, and our teen privacy center has additional resources to help teens understand and manage their privacy across our apps. We are always working on more ways to help keep teens safe, provide them with privacy controls and educate them about how our technologies work.

Question 20. **Beyond surveillance advertisements, are there any other algorithmic-based practices being implemented that are particularly detrimental to children? In your opinion, how can this be mitigated?**

At Meta, we use algorithms to personalize one's experience, and to help connect a person with their friends, family and interests. For example, if a teen moves to a new school or community, our algorithms will help teens find people with similar interests in their community, or help them organize community service efforts. We also use algorithms to help us take action on content that may violate our Community Standards or Recommendation Guidelines. And, we use algorithms to help teens have age-appropriate experiences on our apps. This includes using them to work to filter out content that might be too sensitive, preventing unconnected adults from interacting with teens' accounts, and regularly recommending that teens update their privacy settings.

We also limit the types of content teens can see on our platforms. Our content recommendation controls—known as "Sensitive Content Control" on Instagram and "Reduce" on Facebook—make it more difficult for people to come across potentially sensitive content or accounts in places like Search and Explore. 99% of teens who are defaulted into the most restrictive content and recommendations settings globally and in the US are still using this setting a year later.

We want teens to have safe, age-appropriate experiences on our apps, and we want to help parents manage those experiences. That is why in the last eight years we have introduced more than 30 different tools, resources, and features to help parents and teens. For teens, these tools include nudges that remind them when they have been using Instagram for a while or when it is late and they might want to go to sleep, and the ability to hide words, topics, or people from their experience without those people finding out.

Question 21. **Are you aware of any surveillance advertisements or algorithms that are used to target children, specifically to promote drugs and the sale of narcotics?**

For more information about our advertising standards, please see the responses to your Questions 3 and 6.

Question 22. **Putting aside the debate as to whether use of encryption on a social media platform is good or bad, I'd like to address the point that the use of end-to-end encryption completely blocks the ability for inappropriate images to be detected without user intervention.**

Is it not true, that on a platform such as yours (e.g., Facebook), where you control the application both for sending and receiving information, that interception of inappropriate images can occur when a user would select to send such an image and when a user would receive and view said image?

Our approach to safety in encrypted environments is focused on three key elements: (i) preventing potential harm in the first place; (ii) giving people ways to control their experience; and (iii) responding to violations of our policies quickly. This approach is detailed in our whitepaper, Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging. In an end-to-end encrypted environment, we use machine learning to proactively detect accounts engaged in potentially malicious patterns of behavior. Our machine learning technology will look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect potentially suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors. To help us respond to violations of our policies quickly, we also encourage people to report messages to us in both encrypted and unencrypted services.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people

from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

The technology we explore and develop is designed in accordance with our Security Design Principles. For this reason we have not adopted, and do not intend to develop, scanning technologies that automatically access and report messaging content in end-to-end encrypted messages, often called "client-side scanning." These types of technologies, whether on a person's device or otherwise, without that person's consent and control could be abused by online criminals, malicious hackers or authoritarian regimes, putting people's safety at risk. We do not believe such technologies can be developed and implemented in a manner that is rights-respecting, nor can such technologies meet the expectations people have of end-to-end encrypted messaging services, and significant security concerns have been raised by leading technical experts in the field.

We have spent more than a decade developing policies and technologies to help keep young people safe and to keep predators from attempting to use our service to connect with one another. Our comprehensive approach to safety includes policies prohibiting child exploitation; cutting-edge technology to prevent, detect, remove, and report policy violations as appropriate; and the provision of resources and support to victims. We work with professionals, collaborate

with industry and support law enforcement around the world to fight the online exploitation of children. For example, we respond to valid law enforcement requests for information with data, including email addresses and phone numbers, and traffic data, like IP addresses, that can be used in criminal investigations.

We expect to continue providing more reports to law enforcement than our peers, thanks to our industry-leading work on keeping people safe. For example, WhatsApp—which has long been encrypted—takes action against hundreds of thousands of accounts every month for suspected child exploitative imagery sharing. In 2022, WhatsApp also made over one million reports to NCMEC, all without breaking encryption. This was significantly more than all other encrypted messaging services combined. NCMEC has acknowledged Meta continues to be an industry leader in this work and that Meta "goes above and beyond to make sure that there are no portions of their network where this type of activity occurs."

End-to-end encryption is already widely used by other large messaging services to protect people's private messages and provide people with the privacy and security they expect when messaging friends and family. We do not believe moving to an encrypted messaging environment means sacrificing safety. That is why we will continue to support encryption, while putting features in place to help keep people safe. We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

Questions from Senator Welch

Throughout 2023, Meta carried out multiple rounds of layoffs as part of a plan to eliminate over 10,000 roles. Many of these roles came from crucial teams that ensure the integrity of social media platforms and the safety of users. The reduction of Trust & Safety teams increases the likelihood that harmful content will show up in children's social media feeds and increases the risk of disinformation throughout Meta platforms.

Question 1. **Since Facebook acquired Instagram in 2012, on a quarterly basis, how many Meta employees held positions whose primary responsibility was to study the potential negative impacts of Facebook and/or Instagram on platform users under the age of 18? Please specify how many employees per platform (Facebook and Instagram).**

Meta conducts a variety of user surveys and other research to understand the experiences people have on our platforms. Because of the multi-faceted nature of this work, we do not track information in the manner requested.

After years of growth, Meta implemented a company-wide restructuring plan focused on flattening our organization. The goal of these efforts was to make the company faster, leaner, and more efficient. To be clear, these restructuring efforts did not change the commitment we have to our ongoing integrity efforts. We have brought teams together to think across a number of key issues. For example, our global operations team now works more closely with our integrity team, and we have consolidated certain support teams from different areas across the company.

To be clear, we absolutely remain committed to our work keeping people safe on our services. Even with the targeted changes, we continue to have about 40,000 people focused on overall safety and security efforts. Finding efficiencies in our work has been a focus for years. We will continue to hire across security and integrity teams to support our industry-leading work in the most efficient and effective manner possible.

Question 2. **How many full-time employees do you have working on your Trust and Safety teams today?**

Please see the response to your Question 1.

Question 3. **How many full-time employees did you have on your Trust & Safety teams in 2020?**

Please see the response to your Question 1.

***Question 4*. In 2023 when there were thousands of layoffs at Meta, did content that violates your community policies increase or decrease on your platforms?**

For information on Meta's restructuring, please refer to the response to your Question 1.

We do not track information in the manner requested because there are many factors that contribute to increases or decreases in the amount of violative content reported on platforms, including improvements in detection. We track prevalence, which tells us how often content that violates our standards is seen relative to the total amount of times any content is seen on Facebook or Instagram. We publish the Community Standard Enforcement Report to more effectively track our progress and demonstrate our continued commitment to making Facebook and Instagram safe and inclusive. Updated metrics and recent trends—organized by policy area and extending back over five years—are available at the Meta Transparency Center.

***Question 5.* How much of your content moderation is managed my artificial intelligence?**

Earlier this year, after years of growth, Meta implemented a company-wide restructuring plan focused on flattening our organization. The goal of these efforts is to make the company faster, leaner, and more efficient. To be clear, these restructuring efforts do not change the commitment we have to our ongoing integrity efforts. We have brought teams together to think across a number of key issues. For example, our Global Operations team now works more closely with our integrity team, and we have consolidated certain support teams from different areas across the company.

We have invested more than $20 billion in our overall integrity efforts since 2016 and currently have around 40,000 people working on safety and security. Our investments have allowed us to build technologies to help proactively identify potentially violating content, prioritize critical content for review, and act on content that violates our policies.

We enforce our policies through a combination of people and technology that work to identify violations of our Community Standards across the billions of pieces of content that are posted to our services every day. Technology-driven resources help us identify and take action against violating content and accounts at scale, and assist us in enqueuing certain content for human review. For example, our systems flag content that may violate our policies, people who use our apps report content to us they believe is questionable, and our own teams review certain content. We work to remove content that violates our policies quickly and at scale. We have also built a parallel content review system to flag posts that may be going viral—no matter what type of content it is—as an additional safety net. This helps us catch content that our traditional systems may not pick up. We use this tool to detect and review Facebook and Instagram posts that were likely to go viral and take action if that content violated our policies.

When we began reviewing content on Facebook over a decade ago, our system relied on people to report things they saw and thought were inappropriate. Our teams then reviewed and removed individual pieces of content if they broke our rules. A lot has changed since then. Today, our artificial intelligence (AI) has improved to the point that it can detect violations across a wide variety of areas without relying on people to report, often with greater accuracy than reports from humans. This helps us detect harmful content and prevent it from being seen by hundreds or thousands of people. Further, instead of simply looking at reported content in chronological order, our AI prioritizes the most critical content to be reviewed, whether it was reported to us or detected by our proactive systems. This ranking system prioritizes the content that is most harmful based on multiple factors such as virality, severity of harm, and likelihood of violation. In an instance where our systems are near-certain that content is breaking our rules, it may remove it. Where there is less certainty, it will prioritize the content for teams to review. Technology has also helped scale the work of our content reviewers in areas where there may be a higher frequency of violations. By using technology to help in content determinations, our reviewers can focus on determinations where more expertise is needed to understand context and nuance of a particular situation.

AI has helped to advance our content review process and greatly improved our ability to moderate content at scale. But there are still areas where human review is critical. For example, some determinations, such as whether someone is the target of bullying, require an understanding of nuance and context. Human review is helpful in those instances. And AI relies on training data from reviews done by our teams to identify relevant patterns of behavior and find potentially violating content.

That is why our content review system needs both people and technology to be successful. Our teams focus on cases where it is essential to have people review and we leverage technology to help us scale our efforts in areas where it can be most effective.

Question 6. **Is it your view that artificial intelligence can replace human judgment in identifying and removing false or harmful content? If not, when is human judgment necessary?**

Please see the response to your Question 5.

Question 7. **How have your Trust & Safety teams been trained on how to handle false or illegal AI-generated content?**

As discussed in response to your Question 5, we enforce our policies through a combination of people and technology that work to identify violations of our Community Standards and our

Community Guidelines across the billions of pieces of content that are posted to our services every day. These Community Standards and Community Guidelines apply to all content posted on our services regardless of how it is created, regardless of whether or not it has been generated using AI. Our content review system needs both people and technology to be successful.

Meta's review teams consist of full-time employees who review content as part of a larger set of responsibilities, as well as content reviewers employed by our partners. They come from different backgrounds, reflect our diverse community and have an array of professional experiences—from veterans to legal specialists to enforcement experts in policy areas such as child safety, hate speech and counterterrorism. Our review teams are global and review content 24/7. We have over 20 sites around the world, where these teams can review content in over 50 languages.

In order to do their job, review teams undergo extensive training to ensure they have a strong grasp on our policies, the rationale behind our policies, and how to apply our policies accurately. Reviewers spend at least 80 hours in training with a live instructor. From there, they have hands-on practice using a facsimile of the review system, so they can apply what they have learned in a simulated environment. After this hands-on learning, reviewers get a report highlighting the areas where they apply our policies consistently and accurately and areas where they need more practice. To ensure they are up-to-speed on the latest information, reviewers receive regular coaching, refresher sessions, and policy updates.

Question 8. **How does Meta plan on addressing the large amount of disinformation that could be spread on its platform during the 2024 election?**

We continue to invest in and support our longstanding efforts to deliver value to the people who use our platforms by slowing the spread of viral misinformation and combating disinformation. Our work in these areas is reflected in our content moderation efforts, misinformation policies, third-party fact-checking program, and efforts to fight coordinated inauthentic behavior.

With respect to misinformation, we partner with approximately 100 fact-checking organizations around the world who rate content in more than 60 languages. In the United States, we partner with 11 fact-checking organizations, seven of which cover content in Spanish. Posts debunked by our independent third-party fact-checking partners appear lower in Facebook's Feed, are filtered out of Explore and Hashtags on Instagram, and are shown lower in Instagram's Feed and the Stories tray. We also overlay a warning screen on top of content deemed to be false. People who try to share the content are notified of the fact-checker's reporting and rating, and they are also notified if content they have shared in the past has since been rated false by a fact-checker; content rated false cannot run as an ad.

When it comes to disinformation, we tackle it through our policies and enforcements against coordinated inauthentic behavior (CIB), which covers coordinated networks that centrally rely on fake accounts to mislead people about who they are and what they are doing to manipulate or corrupt public debate for a strategic goal. We conduct our own independent investigations and enforce against CIB. We do so based on the deceptive behavior we see on our platform, not based on the content they share. Our team focused on disrupting influence operations includes experts across the company, with backgrounds in law enforcement, national security, investigative journalism, cybersecurity, law, internet freedom, human rights, and engineering. Our technical teams continue to build scaled solutions to help detect and prevent these violating behaviors, and we work with civil society organizations, researchers, and governments to strengthen our defenses. We have also improved our detection systems to more effectively identify and block fake accounts, which are the source of a lot of inauthentic activity.

We regularly publish Adversarial Threat Reports, which detail the results of our efforts to combat CIB, as well as other adversarial threats we detect and remove from our platforms. Our Q4 2023 report can be found at https://transparency.fb.com/metasecurity/threat-reporting. We also report on our integrity enforcement progress publicly in our Community Standards Enforcement Report. This report includes metrics on how Meta is performing in preventing and removing content that violates our Community Standards and fake accounts.

But it is not enough to just limit misinformation that people might see. We also connect people to reliable information from trusted experts. We do this through centralized hubs like our Voting Information Center, labels that we attach to certain posts with reliable information from experts, and notifications that we run in people's feeds on both Facebook and Instagram. The Voting Information Center on Facebook and Instagram is designed to provide millions of people with accurate information about voting as well as the tools they need to register. It includes: posts from election authorities with announcements about, and changes to, the voting process; links to state voter registration websites, where applicable; links to every state election authority website; and guidance for military and overseas voters. Our goal is to direct Americans to accurate, authoritative voting information, in consultation with state elections authorities, and in doing so, provide a useful tool to supplement their election-related conversations with family, friends, and other sources of information using our services. We hope to give people accurate information so that they may exercise their right to vote and we work hard to protect the integrity of upcoming elections on our family of apps. The Voting Information Center helps to accomplish both.

Going into 2024, our security efforts include: ongoing threat research into and enforcement against new and known threats/threat actors; sharing threat indicators and insights with industry peers so they too can strengthen their responses to foreign interference and other adversarial threats; sharing our threat research with our industry peers, researchers, policymakers, and the

public in our regular adversarial threat reports; continuous detection and blocking of recidivist attempts to come back; and feeding high-fidelity signals derived from threat research into automated detection systems to help scale the work of our security expert investigators allowing them to focus on the most complicated threats.

We will continue to invest in and improve our processes and tools so we can do our part to protect the integrity of future elections. While each election will bring its own unique set of challenges, we are working diligently to apply the lessons we have learned from previous years to the forthcoming 2024 election, and ensuring teams have the appropriate resources to do so.

Historically, Meta has conducted surveys of Facebook and Instagram users regarding users' negative experiences on or as a result of using these social media platforms, including the "Tracking Reach of Integrity Problems Survey" (Facebook) and "Bad Experiences & Encounters Framework" survey (Instagram). All questions below pertain to the use of Meta user surveys.

Question 9. **Describe how user survey results including the "Tracking Reach of Integrity Problems Survey" and "Bad Experiences & Encounters Framework" survey have informed product changes to your platforms to benefit the health or safety of Facebook and Instagram users under the age of 18.**

We use many different tools to understand how users experience Facebook and Instagram, including user perception surveys, such as the ones referenced in your question. User perception surveys help us understand the types of content people say they have seen on the platform. However, those results do not mean that content broke our rules, or that it was objectively harmful, because responses are personal, subjective, and variable. These surveys also do not define the negative experiences we ask users to tell us about. This is intentional: these surveys are global, and different cultures and even different people in the same cultures have different perceptions of experiences in their lives. As a result, it is hard to draw definitive conclusions based on the type of questions asked, but we do use them to inform strategies to help users cope with difficult moments. We recognize that we do not have all the answers and that gathering others' perspectives is crucial to a well-designed platform. That is why we continue to conduct these user experience surveys today.

We have built numerous tools, features, and resources that help teens have safe, positive experiences. Some of the features launched directly address the concerns raised in user perception surveys, such as:

- Notifications to users based on community feedback, recommending they consider revising or taking down potentially hurtful (but non-violating) comments or content;

- Automatic strong warnings based on our machine learning technology when we detect that people try to post potentially offensive comments;
- Using nudges to encourage more people to pause and reflect before replying to a comment that our systems tells us could be offensive;
- Reminding people to be respectful in Direct Messages when sending a message request.

We have also developed tools designed to give young people more control over their experience on Instagram and our other apps—whether that is control over who can contact them or comment on their posts, or control over the kind of language they want to see. Many of these tools were designed thanks to direct feedback from young people.

We work hard to provide support and controls to reduce potential online harms, and it is important to us at Meta that our services are positive for everyone who uses them. Meta has around 40,000 people overall working on safety and security, and we have invested over $20 billion since 2016. This includes around $5 billion in the last year alone.

Question 10. **Has Meta withdrawn any products based on the results of these surveys?**

Please see the response to your Question 9.

Question 11. **Please describe the product and decision-making process within Meta that led to your decision to withdraw or maintain any products following negative or concerning survey feedback.**

Please see the response to your Question 9.

Question 12. **Has Meta used the results of these surveys to block products from launching because they were not safe for children?**

Safety and integrity are key to the experience people have with our services, and Meta builds its services and continually updates them with safety and integrity in mind. We embed teams focusing specifically on safety and security directly into the teams that design and build our services. And we offer integrity tools, built centrally, to individual services teams to allow them to build in preventative safeguards at the start.

When we think an app or feature has serious safety challenges, we will not launch it. After features launch, we continue to monitor their impact, including by looking at integrity metrics, to ensure the features are best serving our community. For more information on user perception surveys, please see the response to your Question 9.

***Question 13*. Please describe the product and decision-making process within Meta that led to your decision to block or proceed with any product launches following negative or concerning survey feedback.**

Please see the responses to your Questions 9 and 12.

All questions below pertain to the use of end-to-end encryption on Meta platforms.

***Question 14*. Please describe why you chose to implement end-to-end encryption on private messages on Meta platforms and the benefits you see from it.**

As our lives move more and more online, we believe it is critical to preserve a space for private conversations where people can have the freedom to be themselves and share their most personal thoughts with loved ones. End-to-end encryption protects the privacy and many other human rights of billions of people every day. End-to-end encryption keeps people and their personal communications safe from malicious hackers, criminals, and authoritarian regimes. It is already widely used by other large messaging services to protect people's private messages and provide people with the privacy and security they expect when messaging friends and family.

For more information, and to review the security principles that serve as reference points for our private messaging design decisions, visit https://engineering.fb.com/2022/07/28/security/five-security-principles-for-billions-of-messages-across-metas-apps/.

***Question 15*. How do you balance the benefits of encryption with the need for law enforcement to be able to track down wrongdoers on your platform?**

Implementation of encryption does not undercut our commitment to work with law enforcement. We have spent more than a decade developing policies and technology to help keep young people safe and to keep predators from attempting to use our service to connect with one another. Our comprehensive approach includes policies prohibiting child exploitation; cutting-edge technology to prevent, detect, remove, and report policy violations; and the provision of resources and support to victims. We work with professionals, collaborate with industry, and support law enforcement around the world to fight the online exploitation of children.

We have developed a streamlined online process through which we accept and review all legal requests from law enforcement. We expedite requests pertaining to child safety, along with other emergency situations, and we have a team dedicated to engaging with NCMEC, International Centre for Missing & Exploited Children, Interpol, the FBI, and numerous other

local, federal, and international law enforcement organizations and departments to help make sure that they have the information and training needed to make the best use of this process and that we are supporting efforts to improve these processes. If we have reason to believe that a child is in imminent danger, we may proactively report relevant information to law enforcement or NCMEC to help safeguard the child.

In an end-to-end encrypted environment, we also use machine learning to proactively detect accounts engaged in potentially malicious patterns of behavior. Our machine learning technology will look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect potentially suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, such as preventing them from interacting with minors. To help us respond to violations of our policies quickly, we also encourage people to report messages to us in both encrypted and unencrypted services.

We expect to continue providing more reports to law enforcement than our peers, thanks to our industry-leading work on keeping people safe. For example, WhatsApp—which has long been encrypted—removes hundreds of thousands of accounts per month for CSAM violations. In 2022, WhatsApp also made over one million reports to NCMEC, all without breaking encryption. This was significantly more than all other encrypted messaging services combined. The National Center for Missing and Exploited Children (NCMEC) has acknowledged Meta continues to be an industry leader in this work and that Meta "goes above and beyond to make sure that there are no portions of their network where this type of activity occurs." We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

Question 16. **Given Meta's decision to implement end-to-end encryption by default, please explain the steps and processes used by Meta on each of its platforms to identify child sexual abuse material, how you remove it, and how you report it to law enforcement.**

Child exploitation is a horrific crime that we work aggressively to fight on and off our platforms. We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. We have processes in place to remove policy-violating content, regardless of the context or the person's motivation for sharing it. We also have developed aggressive, cutting-edge technology to prevent, find, remove, and report policy violating content. In addition to this technology, we have invested in specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic

investigations to both find and review potentially violating content, accounts and adversarial networks.

As a general matter, we do not share detailed descriptions of how our tools work or our enforcement efforts, which, if revealed, could provide a roadmap to highly-motivated bad actors who seek to evade our detection and NCMEC reports, which would ultimately undermine our efforts. That said, we are working hard to further augment the measures we have in place to strive to evolve as predatory behaviors and coded language do as well.

For example, we work to promptly disable accounts on Facebook and Instagram for various violations of our child exploitation policies, such as the apparent malicious[50] distribution of CSAM or sexual solicitation of children. We also recognize that predators may attempt to set up multiple accounts to evade enforcement of our policies. That is why when we disable accounts for these severe violations, we also work to disable explicitly linked accounts (where a person has linked their Facebook and Instagram profiles), high-confidence linked accounts (where we have high confidence that the same person is using multiple accounts), and restrict those devices from setting up future accounts. We also utilize technology and teams to detect and eliminate abusive networks to take on predators who attempt to use our services to connect online. We also collaborate with industry on new programs, such as with Take It Down and Lantern, of which we are founding members. Lantern is a program that enables technology companies to share signals of a child safety threat and Take It Down is a platform designed to proactively prevent young people's intimate images from spreading online. Importantly, we want teens to have safe, age-appropriate experiences on our apps, including on our encrypted services. We do not believe moving to an encrypted messaging environment means sacrificing safety. It is already widely used by other large messaging services to protect people's private messages and provide people with the privacy and security they expect when messaging friends and family. That is why we will continue to support encryption, while putting features in place to help keep people safe.

With respect to Messenger and Instagram Direct Messages, as we rollout end-to-end encryption our approach to safety is focused on three key elements: (i) preventing potential harm in the first place; (ii) giving people ways to control their experience; and (iii) responding to violations of our policies quickly. This approach is detailed in our whitepaper, Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging.

[50] We distinguish between "malicious" and "nonmalicious." In the "malicious" group are people we believe intended to harm children with their content, and in the "nonmalicious" group are people we believe, based on contextual clues and other behaviors, likely did not intend to cause harm to children. (For example, they shared the content with an expression of abhorrence.) Regardless, and consistent with US law, these actors are reported to NCMEC.

To address potential harm, we have built tools and policies specifically to help young people manage interactions with adults. For example, we automatically set teens' accounts under 16 (and under 18 in certain countries) to private when they join Instagram. Private accounts, available to both teens and adults, also prevent other people from seeing friend/follow lists. We also do not allow people who teens do not follow to tag or mention them, or to include their content in Reels Remixes or Guides by default.

We restrict adults over 19 from initiating private messaging with teens who do not follow them on Instagram and with unconnected teens on Messenger. We limit the type and number of direct messages someone can send to an account that does not follow them on Instagram, restricting them to one text-only message until the recipient accepts their request to chat. This helps prevent people from receiving unwanted images, videos, or repeated messages from people they do not know.

We also announced that we plan to introduce stricter default message settings for teens under 16 (and under 18 in certain countries). This means that, by default, teens cannot be messaged or added to group chats by anyone they do not follow or are not connected to on Instagram and Messenger. This is designed to help teens feel even more confident that they will not hear from people they do not know—including other teens—in their private messages.

We also recently announced that we are testing our new nudity protection feature in Instagram direct messages, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to help protect people from seeing unwanted nudity in their direct messages, but also to help protect them from scammers who may send nude images to trick people into sending their own images in return. Nudity protection uses on-device machine learning to analyze whether an image sent in a direct message on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta will not have access to these images—unless someone chooses to report them to us.

Nudity protection will be turned on by default for teens under 18 globally, and we will show a notification to adults encouraging them to turn it on. When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they have changed their mind. Anyone who tries to forward a nude image they have received will see a message encouraging them to reconsider. When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient is not confronted with a nude image and they can choose whether or not to view it. We will also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat. When sending or receiving these images, people will be directed to safety tips, developed with

guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they are not who they say they are. These safety tips also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.

In addition, in an end-to-end encrypted environment, we use machine learning to proactively detect accounts engaged in malicious patterns of behavior. Our machine learning technology will look across non-encrypted parts of our platforms—like account information and photos uploaded to public spaces—to detect suspicious activity and abuse. For example, if an adult repeatedly sets up new profiles on Facebook and Instagram and tries to connect with minors they do not know or messages a large number of strangers, we can intervene to take action, including preventing them from interacting with minors.

We have also built more than 50 tools, resources, and features to help support teens. For more information on these tools as well as to review resources from experts, visit our Family Center: https://familycenter.meta.com/.

To help us respond to violations of our policies quickly, we encourage people to report messages to us in both encrypted and unencrypted services. We have made our reporting tools easier to find and started encouraging teens to report at relevant moments, such as when they block someone. On Instagram, we have developed proactive prompts—or safety notices—that notify young people when an account that has been exhibiting potentially suspicious behavior is interacting with them in messages and give teens an option to block, report, or restrict the account. We also developed safety notices in Messenger, which provide tips on spotting potentially suspicious activity and educating people on how to take action. These notices help people avoid scams, spot impersonations, and flag accounts that have been exhibiting potentially suspicious behavior that attempt to connect to minors.

Keeping young people safe online has been a challenge since the advent of the internet. Online predators are determined criminals who use multiple apps and websites to target young people. They also test each platform's defenses, and they learn to quickly adapt. That is why now, as much as ever, we are working hard to stay ahead of these threats by developing technology to root out predators, work with specialists dedicated to online child safety, and share information with our industry peers and law enforcement.

Questions from Senator Whitehouse

Question 1. **What exemptions from the protections of Section 230 would your company be willing to accept?**

While we at Meta are working to make progress, we know that we can not—and should not—do it alone. That is why we support updated regulations to set clear and fair rules, and support a safe and secure open internet where creativity and competition can thrive. The last time the United States enacted comprehensive internet regulation was in 1996 when updates to the Communications Act closed a major gap in liability issues for online content by creating Section 230. Technology has evolved exponentially in the last quarter-century, and the rules should keep pace.

Meta has long been supportive of updating Section 230, for example, to ensure that it separates good actors from bad, by making sure that companies cannot hide behind Section 230 to avoid responsibility for intentionally facilitating illegal activity on their services. We understand that people want to know that companies are taking responsibility for combating harmful content—especially illegal activity—on their online services. They want to know that when such services remove content, they are doing so fairly and transparently.

In addition to concerns about unlawful content, Congress should act to bring more transparency, accountability, and oversight to the processes by which companies make and enforce their rules about content that is harmful but legal. While this approach would not provide a clear answer to where to draw the line on difficult questions of harmful content, it would improve trust in and accountability of the systems and address concerns about the opacity of process and decision-making within companies. This is why we agree that online services should strive toward enhancing transparency.

Updating Section 230 is a significant decision. It is important that any changes to the law do not prevent new companies or businesses from being built, because innovation in the internet sector brings real benefits to all Americans, as well as to billions of people around the world. We stand ready to work with Congress on what regulation could look like, whether that means Section 230 reform or providing guidance to services on other issues such as harmful content, privacy, elections, and data portability. By updating rules for the internet, we can preserve what is best about it—the ability for people to express themselves and for entrepreneurs to build new things—while also protecting society from broader harms.

Question 2. **Is it your belief that your company should enjoy absolute immunity under Section 230 from suits like Doe v. Twitter, No. 21-CV-00485-JCS, 2023 WL 8568911 (N.D.**

Cal. Dec. 11, 2023), no matter the extent of your company's failure to remove reported child sexual abuse material from the platform or to stop its distribution?

Child exploitation is a horrific crime that we work aggressively to fight on and off our services. We have spent more than a decade developing policies and technology to help keep young people safe and to keep predators from attempting to use our service to connect with one another. We have strict policies against child nudity, abuse, and exploitation, including child sexual abuse material (CSAM), inappropriate interactions with children, solicitation, exploitative intimate imagery and sextortion, and content that sexualizes children. We have processes in place to remove policy-violating content, regardless of the context or the person's motivation for sharing it. We also have developed aggressive, cutting-edge technology to prevent, find, remove, and report policy violating content. In addition to this technology, we have invested in specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations to both find and review potentially violating content, accounts and adversarial networks.

Thanks to these efforts, we find and report more CSAM to the National Center for Missing and Exploited Children (NCMEC) than any other service today. In 2022, all of the industry made 32 million reports to NCMEC collectively. We made over 26 million reports between Facebook and Instagram. The rest of the industry made less than 6 million reports collectively. NCMEC has acknowledged Meta as an industry leader in this work, as have other child safety organizations. We are committed to leading in the fight against online child exploitation and will continue to devote significant resources and attention to these efforts.

As discussed in response to your Question 1, Meta supports thoughtful reform of Section 230. People of all political persuasions want large companies to take responsibility for combating illegal content and activity on their services. And when they remove harmful content, people want them to do so fairly and transparently. The sheer volume of user-generated content on the internet means that online companies have to make decisions about how to organize, prioritize, and deprioritize this content in ways that are useful to people and advertisers, while enforcing our policies against harmful content.

Services should be granted continued protection from liability for the content they carry if they can demonstrate that they have robust practices for identifying illegal content and quickly removing it. While it would be impractical to hold companies liable if a particular piece of content evades detection, they should be required to follow common industry standards and best practices.

FIGHTING HATE FOR GOOD

Chairman Dick Durbin
Ranking Member Lindsey Graham
Senate Judiciary Committee
United States Senate
224 Dirksen Senate Office Building
Washington, DC 20510

February 2nd, 2024

Dear Chair Durbin, Ranking Member Graham, and members of the Judiciary Committee,

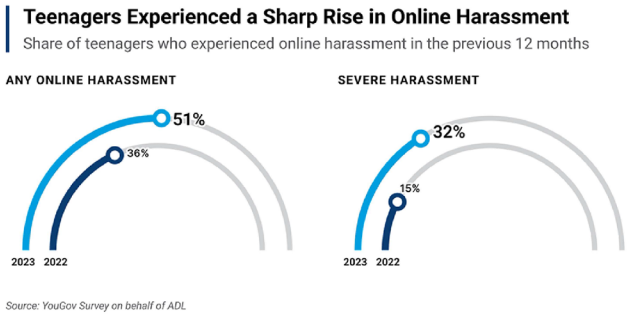
As you know, ADL (Anti-Defamation League) has been a leader in the fight against hate and antisemitism for over a century, rooted in and drawing upon the lived experience of a community relentlessly targeted by extremists, bigots, and other bad actors. Since 2017, the ADL Center for Tech and Society (CTS) has provided unique expertise because of ADL's work at the intersection of civil rights, extremism, and tech.

We are grateful that the Senate Judiciary Committee continues to explore avenues to curb the ongoing proliferation of online child sexual abuse material (CSAM), as platforms continue to struggle to rein in or even prove that they are slowing down circulation.[1] In this tenuous moment, the explosion of generative artificial intelligence (GAI) looms as an accelerant for one of the oldest and most intractable problems created by the internet.[2] We appreciate the committee's effort to address online CSAM by considering a suite of bills including the STOP CSAM Act, the EARN IT Act, the SHIELD Act, the Project Safe Childhood Act, and the REPORT Act. However, these bills and others intended to protect children online should go beyond CSAM to also include the pervasive hate and harassment faced by these platforms' most vulnerable consumers. Among the research initiatives at CTS is the Online Hate & Harassment (OHH) Survey, a nationally representative annual survey to understand how many American adults experience hate or harassment on social media. The 2022 and 2023 editions of the OHH survey also included nationally representative samples of teenagers aged 13-17, positioning CTS well to comment on platforms' inability or unwillingness to protect young adults from identity-based hate and harassment.[3]

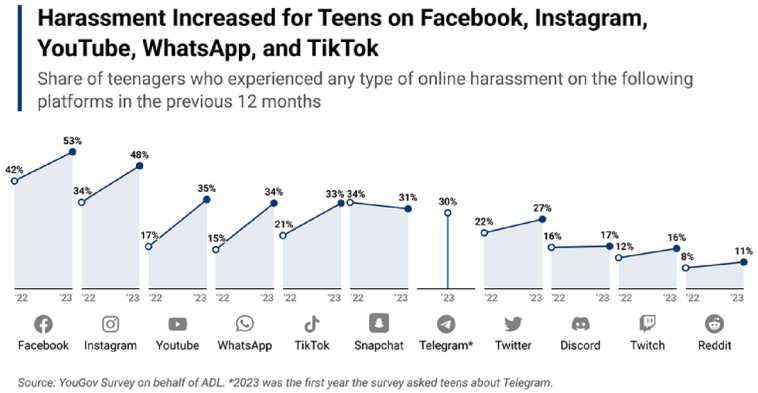
In the 2023 edition of the OHH Survey, CTS found that 51% of the 550 teenagers interviewed experienced some form of online harassment in the preceding twelve months—a 15 percentage point increase from the 2022 edition of the survey, and an increase 5 points greater than the year-over-year

[1] Washington Post article "Child sex images are booming online. Congress wants to know why." (January 2024) https://www.washingtonpost.com/technology/2024/01/28/csam-ncmec-senate-hearing-child-porn/
[2] Stanford Internet Observatory study "Identifying and Eliminating CSAM in Generative ML Training Data and Models" (December 2023) https://purl.stanford.edu/kh752sm9123
[3] CTS report "Online Hate and Harassment: The American Experience 2023" (June 2023) https://www.adl.org/sites/default/files/pdfs/2023-12/Online-Hate-and-Harassmen-2023_0_0.pdf

increase reported by adults. 32% of teenagers surveyed also reported facing severe harassment, a 17-percentage point increase from the 2022 results.



The graphic below, reproduced from the survey, demonstrates that when results were sorted by platform, harassment of teenagers increased from 2022 to 2023 on almost every entity that recently appeared before the Committee.



Overall, reports of each type of hate and harassment increased by nearly every measure and within almost every demographic group. Among the most troubling insights we derived specifically from the teenage respondents was that 47% who indicated they were harassed online in the past 12 months reported subsequent online/in-person harassment.

While identity-based harassment of teenagers is not as gut-wrenching as CSAM exploiting young children, our data demonstrate that any discussion of platform safety and the dangers of being online for

young people is incomplete without discussing the pervasiveness of severe harassment. Meta Platforms, for example, is being sued by multiple attorneys general for knowingly designing and deploying features that harm young users' mental health.[4] As such, we urge all members of the committee to redouble their efforts to rein in big tech while also maintaining a wide view of the multidimensional nature of platforms' adverse impact on young users. ADL has endorsed the bipartisan Platform Accountability and Transparency Act (PATA) re-introduced in the 118th Congress by Senator Coons. The transparency provisions of PATA would provide a data-driven understanding of the effects platforms are having on young users' well-being.

ADL is appreciative of the committee's leadership on these complex and important issues, indicative of the bipartisan commitment to protecting children online. We are glad to see the committee using whatever means necessary to compel testimony from these platforms as they continue to rake in record profits while failing to mitigate the dangers their products pose to vulnerable users. ADL looks forward to continued work on these issues with the committee.

Sincerely,

Max Sevillia

Senior Vice President, National Affairs

Anti-Defamation League

[4] The Hill article "Zuckerberg 'vetoed,' 'ignored' plans to boost teen well-being on Meta platforms, lawsuit alleges" (Nov 2023) https://thehill.com/policy/technology/4302145-zuckerberg-vetoed-teen-mental-health-facebook-instagram

Alliance to End Human Trafficking
(AEHT)

Statement for the Record

U.S. Senate on the Judiciary
Full Committee Hearing
Protecting Children Online

January 31, 2024

Submitted by:

Katie Boller Gosewisch
Executive Director

Physical Location:
7575 Grand River Avenue
Suite 109
Brighton, MI 48114

Mailing Address:
P.O. Box 713492
Chicago, IL 60677-4392

Office: 267-332-7768
Mobile: 763-567-1086

alliancetoendhumantrafficking.org

The Alliance to End Human Trafficking (AEHT) commends the Senate Judiciary Committee for the public dialog on the important issue of protecting children online. We thank the Committee for addressing this vitally important issue.

AEHT is a collaborative, faith-based national network that offers education, supports access to survivor services, and engages in advocacy to eradicate human trafficking. Our efforts focus on educating the public, preventing any infringement on human dignity, and aiding survivors in leading fulfilling lives. Our members include over 220 congregations of Catholic Sisters, coalitions working against human trafficking, and individuals who share our mission across the United States. Additionally, we are the U.S. member of Talitha Kum, an international organization dedicated to ending human trafficking.

Congress is correctly poised to pass the pending bipartisan legislation[1] aimed at protecting children online. This legislation can pave the way for harnessing technology for beneficial purposes, such as academic learning. Simultaneously, it aims to safeguard children from potential vulnerabilities and sexually abusive grooming through technological platforms, tools, and applications that target them and facilitate online connections with nefarious actors, resulting in detrimental and devastating outcomes, including human trafficking.

AEHT is also poised to help inform the Committee's work on this important issue. Our members have been working for more than a decade to end the scourge of human trafficking, which includes online trafficking of children.[2] Online child trafficking does not

[1] The *STOP CSAM Act* supports victims and increases accountability and transparency for online platforms. The *EARN IT Act* removes technological companies' blanket immunity from civil and criminal liability under child sexual abuse material (CSAM) laws and establishes a National Commission on Online Child Sexual Exploitation Prevention; The *SHIELD Act* ensures that federal prosecutors have appropriate and effective tools to address the nonconsensual distribution of sexual imagery. The *Project Safe Childhood Act* modernizes the investigation and prosecution of online child exploitation crimes. The *REPORT Act* combats the rise in online child sexual exploitation by establishing new measures to help strengthen reporting of those crimes to the CyberTipline.

[2] Broadly, human trafficking is a modern-day form of slavery. It is a crime under state, federal and international law. It is currently the second largest type of criminal activity, exceeded only by the illegal drug trade. This crime occurs when a trafficker uses force, fraud or coercion to control another person for the purpose of engaging in commercial sex acts or soliciting labor or services against his/her will. Sex

necessitate physical contact between the child and the trafficker. In fact, internet-based child sex trafficking often empowers exploiters to lure or groom minors into creating explicit content (i.e., self-made Child Sexual Abuse Material (CSAM[3]).

To highlight some of our credentials, we offer the following brief biographies to illustrate the depth of our commitment to protecting children and the extent of our work:

- **Ann Oestreich, IHM, a member of the Sisters, Servants of the Immaculate Heart of Mary** in Monroe, Michigan, has served in education, communications, and social justice ministries. As President Emeritus of AEHT, she has extensive board experience and served as the North American representative on the Talitha Kum International Coordination Committee.[4] Her leadership helped establish and grow AEHT's work and commitment to the vulnerable, including protecting children online.

- **Jeanne Christensen, RSM, is a member of the Sisters of Mercy of the Americas,** for more than sixty years. She serves as their Justice Advocate Against Human Trafficking. She has been a member of AEHT since its founding. She has led the work of AEHT by establishing our advocacy initiatives which promote education, awareness, and prevention of human trafficking. Additionally, she led a group of our members who are focused on local, state, and federal development to address the root causes of human trafficking. She is a frequent speaker and recognized leader with substantial experience and substantive knowledge of human trafficking.

---

trafficking is when a commercial sex act is induced by force, fraud or coercion, or in which the person induced to perform such act is under 18 years of age. Labor trafficking is the recruitment, harboring, transportation, provision or obtaining of a person for labor or services through the use of force, fraud or coercion for the purpose of subjection to involuntary servitude, peonage, debt bondage or slavery. Other forms of trafficking include organ removal, mail-order brides/forced marriages, and child soldiers.

[3] Human Trafficking Front has compiled information on this topic available at: https://humantraffickingfront.org/online-exploitation-children/#:~:text=Online%20child%20sex%20trafficking%20involves,made%20CSAM%20is%20a%20trafficker.

[4] Talitha Kum was established in 2009 with the International Union of Superiors General (UISG) as an international initiative against human trafficking and exploitation. Talitha Kum promotes collaboration among networks organized at national, regional and continental level, actively supporting victims, survivors and people at risk. Consult https://www.talithakum.info/ for more information.

- **Margaret Anne Meyer, MMM, Medical Missionaries of Mary,** has spent over 20 years working to stop human trafficking. She has a sharp focus on combating the victimization of children through her advocacy work at the federal, state, and local levels. In her experience, one can make a difference in ending human trafficking through direct contact with governmental representatives and legislators where she communicates the needs and vulnerabilities of children. She actively monitors federal legislation that involves online safety of children. She is a steadfast advocate for keeping children safe from predators.

- **Maryann Agnes Mueller, CSSF, is a Felician Sister of North America.** She serves as the full-time justice and peace coordinator in Enfield, Connecticut. She publishes AEHT's monthly "Stop Trafficking" Newsletter. The most-recent issue addresses the grooming of children on social media platforms that is referenced herein and attached in the Appendix for your reference.

- **Pat Millen, OSF, with the Sisters of Saint Francis of Philadelphia,** has been in religious life for over 46 years. Her ministries include serving as a court-appointed special advocate guardian ad litem for children in Washington State. She has experienced first-hand the consequences of human trafficking trauma and its devastating effects on children. She is the Justice, Peace, and Integrity of Creation (JPIC) Coordinator for the West.

These "snapshots" represent the long-standing commitment of women religious in their ministries to be a voice for the voiceless. AEHT has many members and staff who work on a daily basis with the goal of protecting children and others from human traffickers, including in online environments.

The Current Landscape

Children are uniquely vulnerable to human trafficking and exploitation, and oftentimes become unwitting victims of trauma, physical, sexual, and mental abuse, among other negative impacts. Human trafficking removes individuals' human dignity, reducing them to commodities that can be bought and sold. It steals the innocence of children, causing lasting damage to their lives and making it challenging for them to regain a sense of self-worth and value.

Human traffickers exploit the vulnerabilities of others and unfortunately, online platforms, including social media sites, can facilitate predators' access to and exploitation of children and teens. In the online world, children's vulnerabilities can include low self-esteem, isolation, a present or past history of abuse, a desire for love or friendship, poverty, seeking work, the pursuit of a dream, trying to belong to a group or to "fit-in," and many other factors. Deception, empty promises, threats, and manipulation by traffickers occur online unbeknownst to parents as children are swept into situations they cannot anticipate and are not equipped to handle.

In the contemporary era, children seamlessly integrate technology into their daily routines, engaging in a spectrum of activities that include early-childhood lessons for learning numbers and colors, interactive online gaming, educational lessons, homework help, collaborative group assignments, social interactions via social media and gaming, presentations, and both recorded and live-streamed videos. The multifaceted utilization of technology through online platforms, tools, and applications begins at a young age and has become an integral part of the American way of life.

During the COVID-19 pandemic, technology often served as a primary modality used for learning. It was also the one outlet from social isolation to foster human connection that was not possible in an in-person way. As the country adopted technology and became more accustomed to its scope and uses, the post-pandemic years have continued to leverage technology. We expect even broader uses and applications to come.

Given this reality, **it is imperative that Congress take action to ensure that children are protected online from being trafficked.**

Statistics and Research

Various agencies within the U.S. government have identified the risks children face online and the means and methods for human trafficking. The Department of State issues an annual Trafficking in Persons Report[5], which provides both a domestic and international perspective on human trafficking. Migration and international schemes contribute to this largely underreported crime.

The Polaris Project (Polaris) is nationally known for examining data from the National Human Trafficking Hotline. From January 2020 through August 2022, Polaris identified the following conclusions using data from hotline reports[6]: (a) in situations of sex trafficking, 37% of potential victims were minors; and (b) in situations of labor trafficking, 9% of potential victims were minors.

The 2022 Federal Human Trafficking Report (FHTR)[7] used different data to analyze human trafficking methods and trends. The 2022 FHTR identified that "the internet has remained the most common location for recruitment of victims of trafficking for 23 years. Since 2000, the FHTR has identified 11% of sex trafficking victims as being recruited online, primarily through social media platforms, web-based messaging applications, online chat rooms, dating [applications], classified advertisements, or job boards."

In addition, the FHTR notes that "[w]hile the commercial sex industry has existed for centuries, how the industry operates has changed significantly in recent decades with the introduction of technology and the internet. Access to the internet and mobile

[5] https://www.state.gov/reports/2023-trafficking-in-persons-report/
[6] https://polarisproject.org/wp-content/uploads/2020/07/Hotline-Trends-Report-2023.pdf
[7] https://traffickinginstitute.org/wp-content/uploads/2023/06/2022-Federal-Human-Trafficking-Report-WEB Spreads_compressed.pdf

devices allows buyers of commercial sex to purchase commercial sex easily and discreetly through online platforms and social media. In 2022, the primary method traffickers used to solicit buyers of commercial sex was through the internet (53%).

It is also imperative to note the variety of authorities that have documented the online exploitation of children. The National Center on Sexual Exploitation annually compiles a "Dirty Dozen"[8] list to identify highly-used technology company platforms, tools, and applications that present detrimental and substantial risks to children and to explain how they proliferate and enable online child exploitation. In December 2022, the U.S. Government Accountability Office (GAO) issued a report[9] that examined the increasing "volume, complexity, and danger of sexual exploitation of children online . . . [and that addressing] the issue is a formidable task amid a digital landscape characterized by increasing online access, advances in technology, and increased use of encryption." The GAO report recommended federal governmental efforts to address online exploitation of children, including a continual focus on our National Strategy to combat exploitation of children.[10]

Likewise, the U.S. Department of Justice has recognized that "underlying every sexually explicit image or video of a child is abuse, rape, molestation, and/or exploitation," The production of CSAM creates a permanent record of the child's victimization," which in some cases has led to a child's suicide and always leads to lasting psychological and/or physical trauma.[11] This includes livestreaming and virtual child sex trafficking with an expansive customer base through domestic and international access to profits from child victims.[12] Media reports have cited well over 29.3 million CSAM in 2021.[13]

---

[8] https://endsexualexploitation.org/dirtydozenlist-2023/
[9] https://www.gao.gov/assets/gao-23-105260.pdf
[10] *Id.*
[11] https://www.justice.gov/d9/2023-06/child_sexual_abuse_material_2.pdf
[12] https://www.justice.gov/d9/2023-06/livestreaming_and_virtual_child_sex_trafficking_2.pdf
[13] https://www.theguardian.com/technology/2022/mar/24/sites-reported-record-293m-child-abuse-images-in-2021.

Studies have concluded that to counteract online child sexual abuse, it is essential to design effective strategies, since the internet is a hub of online sexual abuse activities.[14] While some have attempted to document the prevalence of sex trafficking of children and adolescents in the United States, it is generally known that much of the exploitation and trafficking go largely undocumented and data can be hard to find, despite efforts by enforcement authorities and nonprofit entities to combat these realities.

AEHT's Ongoing Efforts and Current Focus

AEHT publishes a monthly newsletter, *Stop Trafficking!*,[15] to address issues and trends that predominate and facilitate human trafficking. Our most recent issue is the first of a two-part publication discussing how the online environment can facilitate the grooming of children for sexual exploitation and trafficking. The full text of our publication is attached in the Appendix for your reference. Some of the key points include:

- We recognize that since the beginning of social media, child sexual exploitation has become one of the biggest challenges for tech companies.
- The internet has revolutionized the way traffickers groom and recruit potential victims.
- Traffickers often use covert communication when advertising or selling their victims. These posts carry hidden meanings understood only by those involved in the buying and selling of children.
- Children are especially vulnerable to grooming by traffickers online. Their brains undergo a shift around age 10, which encourages them to seek social rewards, especially attention and peer approval.
- Trafficking can be hard to recognize, especially for those being groomed or trafficked.

[14] https://www.researchgate.net/profile/Sana-Ali-13publication/352120820_Child_Sexual_Abuse_and_the_Internet-A_Systematic_Review/links/63d8ef10c465a873a271c159/Child-Sexual-Abuse-and-the-Internet-A-Systematic-Review.pdf?_sg%5B0%5D=started_experiment_milestone&_sg%5B1%5D=started_experiment_milestone&origin=journalDetail&_rtd=e30%3D

[15] The January 2024 issue is available at: https://alliancetoendhumantrafficking.org/wp-content/uploads/2024/01/2024_1-Stop-Trafficking-EN.pdf

- Livestreaming child sexual exploitation (LCSE) routinely takes place online. There are three types of LCSE:

    - **Child self-generated**, which occurs when a trafficker coerces a child to engage in sexually-explicit conduct on a live stream, usually from a child's bedroom or bathroom.
    - **Offender streaming**, which occurs when a perpetrator, often a family member or friend, sexually abuses a child in person while live-streaming the abuse to viewers. The viewers typically do not know each other, and they often participate by requesting that specific sex acts be committed.
    - **Virtual child sex trafficking,** which is when offenders pay to watch while another offender sexually abuses a child in person, or offenders pay a victim directly to create self-generated CSAM.

Our Recommendations for How Congress Can Act Now to Protect Children

AEHT supports each of the bipartisan bills to help stop the exploitation of children online. While technology can serve as a helpful tool for accessing information, the dangers for children are too complex and too risky for them to navigate alone. **Congress can and should enact federal requirements for technology companies to protect children from human trafficking and to be held accountable when trafficking and other exploitation occurs as facilitated by the platforms, tools, and applications they offer.**

We encourage the Committee to remain steadfast in maintaining key components of the legislation and to consider our recommendations for its legislative constructs:

(1) Technology companies should be required to do more to prevent, detect, and report human trafficking. Despite the known risk to children, technology companies have failed to sufficiently expose, combat, and stop CSAM and human trafficking on their platforms, tools, and applications. Despite public

accounts from victims and survivors, governmental and private research and studies, media reports, and documented cases of exploitation and trafficking, technology companies persist in conducting business as usual. Although companies are aware of the ongoing instances of abuse and trafficking taking place on their platforms and applications, there has been no sense of urgency to-date to change their business practices and processes. The time is ripe for new and urgent responses by these companies to stop enabling abuse and trafficking of our Nation's youth.

(2) **Technology companies cannot and should not be immune from liability when children are harmed through the products, platforms, tools, and applications they offer.** For an extended period, technology companies have benefited from the "section 230"[16] immunity, originally implemented to foster the Internet's development and to facilitate the exchange of information without holding companies accountable for content and actions on their platforms, tools, and applications. However, the days of the "early Internet" are over and technology companies can no longer claim that they do not have the resources to combat CSAM and human trafficking that their platforms, tools, and applications facilitate. It is time for a fresh approach to stimulate technology companies to act to protect children online.

A Congressional Research Report[17] noted that "[c}ourts have interpreted Section 230[18] to foreclose a wide variety of lawsuits and to preempt laws that would make providers and users liable for third-party content. For example, the law has been applied to protect online service providers like social media companies from lawsuits based on their decisions to transmit or take down user-generated content." This is crucial to note because it is recognized that technology companies do, in fact, possess the ability to and routinely do take down

[16] *Id.* at fn 6.
[17] https://crsreports.congress.gov/product/pdf/R/R46751
[18] 47 U.S.C. 230. Section 230 was enacted as part of the Communications Decency Act of 1996 and amended the Communications Act of 1934. This provision provides federal immunity to technology companies for information provided by another person (i.e., a technology user).

user-generated content. Despite this reality, CSAM and human trafficking still proliferate online. **Congress should remove the long-standing immunity protection that technology companies enjoy while child victims experience harm and trauma without recourse against the companies that facilitated their abuse and trafficking.**

(3) **Algorithms that are used to direct content to children should be transparent and made public. Proactive detection algorithms that analyze content and online behavior should be used to identify CSAM and potential online trafficking.** Knowing how technology companies direct traffic, content, advertisements, and make online connections to children should be readily accessible for parents, guardians, educators, and others who need this information to evaluate the quality of protections available by platform. By making algorithms publicly available, consistency and best practices can be promoted across platforms as they continue to strive to keep pace with changing patterns and trends that evolve and place children at risk.

(4) **Children should be able to restore their dignity and seek compensation from those who facilitate or are directly responsible for the harm they suffer. This includes having the right to demand that CSAM be removed from platforms and applications and not be subject to continuing distribution.** AEHT believes that children and their parents or guardians should be able to seek recourse from technology companies when they are harmed. Likewise, children and their guardians should have the right to request – and to require the technology companies to act within a mandated time frame – to remove CSAM from platforms and applications. **While guidelines can allow for successful prosecution of the individuals involved, we encourage Congress to establish a process for requests to remove CSAM and timeframes for doing so.**

(5) **Technology companies should be required to conduct independent audits or to deploy tools that identify potential CSAM or instances of suspected human trafficking. When identified, reports to law enforcement should be made.** Technology companies have become accustomed to policing certain activities that take place on their platforms, tools, and applications, (e.g., "fact checks," "community notes," de-platforming users, "shadow banning" users) while allowing CSAM and trafficking to be largely unchecked and eliminated. The ability exists for technology companies to do more to stop CSAM and human trafficking.

(6) **Encrypted technologies and messaging should not be available to minors.** As technology evolves, the use of encryption is becoming more commonplace. This allows communications to take place via a confidential and unbreakable stream. AEHT members continue to express concern that the adoption of these encryption technologies may be embraced by abusers and traffickers, providing them the means to broaden their predatory activities targeting children without any means to observe, prevent, or detect CSAM, abuse, and trafficking. Therefore, minors should not have the capacity to engage with encrypted technologies in online environments.

(7) **Parents should have the ability to approve with whom children are connecting online.** Technology companies should develop processes to permit parental oversight.

(8) **Victim-centered processes and procedures should be implemented to help survivors of human trafficking. Technology personnel must comprehend the impact on children and adults who are subjects of CSAM and those who have been trafficked.** Such understanding is necessary for these professionals to grasp the repercussions of their business practices and policies on individuals. Only with such perspectives can survivors and companies collaborate to address this horrific problem. Moreover, technology companies

should be knowledgeable about referring victims to appropriate resources to aid in their recovery and healing from abuse and trafficking.

(9) **Steps should be taken by technology companies to work with law enforcement.** Through accountability and public exposure of bad actors and their methods, only then can deterrence begin.

(10) **Increased funding should be available for the National Human Trafficking Hotline and other reporting mechanisms.**

Countering the Arguments Against Protecting Children

The technology companies have offered a three-pronged argument for opposing federal legislation on this topic. Essentially, their positions to date are that: (1) individual's freedom of speech would be negatively impacted; (2) individual's privacy would be compromised; and (3) encryption technologies cannot be broken and thus they cannot be held responsible for information that is encrypted. Our responses below summarize why these arguments are misleading.

Individual Freedom of Speech Will Not Be Negatively Impacted

It is perplexing that technology firms would argue that safeguarding children online through impending legislation would compromise online freedom of speech. This assertion distorts the fundamental principles of our First Amendment rights. While the protection of freedom of speech is vital, it is not without limits. Legal precedent has already established boundaries on the types of speech that warrant protection. The audacity of technology companies in making such a claim is truly astonishing.

With political preferences aside, it is well-known that technology companies "deplatformed" and continue to either censor, amend through "community notes," or

"shadowban" certain opinions and voices on their platforms. It is well-known that the then-popular Parler application was removed from users' availability following the 2020 election; and many voices with differing opinions or real-life experiences about COVID-19, governmental lockdown strategies during the pandemic, and many other topics have been either suppressed or deplatformed by the technology companies as disinformation or misinformation. While we offer no opinion on the validity of any of the political opinions expressed or that wish to be expressed online, the mere fact that technology companies *already police content expressed on their platforms, tools, and applications* obliterates their assertion that they are somehow concerned about free speech. More importantly, there is no legal or moral precedent that remotely provides protection for criminal acts involving CSAM, human trafficking, and online abuse of children. Therefore, freedom of speech has nothing to do with protecting innocent children online.

Individual's Privacy Will Not Be Compromised

Technology companies have asserted that online privacy can be negatively affected if they are required to police images and content that is shared online. This is a misnomer; as noted above, technology companies have the technological ability and tools to scan and detect CSAM and potential human trafficking. No adult possesses a right to privacy for creating, distributing, possessing, engaging in, or profiting from CSAM and human trafficking.

Encryption Technologies Exist and Should Be Appropriately Dealt with by Technology Companies

We noted above our specific recommendations for federal legislation and encryption, which include barring children from encrypted messaging solutions without parental consent. The current technological landscape already offers a number of encrypted messaging options, and technology companies have been paving the way for encryption as a way to protect online data for some time. However, encryption solutions

should not be deployed online without careful consideration about the negative consequences that can result for our Nation's youth. Nefarious actors, pedophiles, and traffickers can avail themselves of encryption to hide CSAM and human trafficking. Guardrails should be put in place by Congress to protect children from a technology tool that can be used to exploit and traffick them.

Alliance to End Human Trafficking
(AEHT)

Appendix

Stop Trafficking!
AwarenessAdvocacyAction
Anti-Human Trafficking Newsletter • January 2024 • Vol. 22 • No. 1

FOCUS: This month's newsletter is the first of two newsletters to focus on the grooming and trafficking of children on social media.

In the years since the beginning of social media, child sexual exploitation has become one of the biggest challenges facing tech companies. The internet has revolutionized the way traffickers groom and recruit potential victims. Every year, more and younger children are given sometimes unmonitored access to devices that connect them to the internet. Children are especially vulnerable to exploitation as their brains prompt them to seek social rewards, including attention and peer approval. Perpetrators who otherwise would not have contact with children can easily and instantaneously connect online with potential victims anywhere in the world.

Recruitment Using Social Media

Traffickers use social media platforms and chat rooms used by children and teens, usually creating fake profiles (catfishing) or pretending to be someone the child knows, to target and groom child victims for sex trafficking.

The process of grooming and recruiting victims online for exploitation involves several stages.

The perpetrators look for children who appear vulnerable by what they post on social media or gaming communities. Research conducted by the United Kingdom's National Society for the Prevention of Cruelty to Children found that children who share feelings of vulnerability on social media are at higher risk of being targeted and groomed by perpetrators online.

The perpetrator initiates contact with the potential victim, usually using a fake identity or profile, and builds a rapport with the child. The perpetrator will create a sense of friendship or love relationship.

Children will sometimes have two accounts on a social media platform; one they show their parents if their parents check their phones, and one they do not.

Traffickers may promise desirable items such as games, toys, gadgets, or clothing to entice the child and strengthen the bond.

The perpetrators understand the child's vulnerabilities, telling them what they want to hear to gain their trust. Some examples of how a trafficker may respond to a child or teen's vulnerability include:

Victim expresses:	**Trafficker's Response:**
Feeling misunderstood	I understand you
Feeling of inadequacy with appearance, etc.	You are beautiful/I am proud of you
Lack of freedom	You are mature for your age and can make your own choices

Awareness

Once gaining their trust, the predator may request or exchange explicit images or videos with children, often starting with innocent conversations and gradually escalating to more explicit content. Once the predator obtains compromising material, they can use it to manipulate and blackmail the child, sometimes called "sextortion," into further exploitation or money from the child by threatening to release the images or information. This type of grooming can occur on virtually any online platform, including Facebook, Instagram, Twitter, Tumblr, Snapchat, TikTok, WhatsApp, Kik, Tinder, and Ask.Fm.

Some traffickers may introduce the child to a group or online community where exploitation is normalized, and other children are already involved. Through manipulation and the desire to fit in, the perpetrator coerces the child into engaging in sexual activities or agreeing to meet offline.

In some cases, traffickers use the grooming tactic of asking children to do "dares" that escalate into sexual activity online. For example, traffickers may begin to dare children to perform sexual behaviors, from removing clothes to increasingly sexual demands to engaging in sexual acts online. This material may then be used to coerce the child to meet the perpetrator in person, threatening to expose the images if they refuse. Once the child and the perpetrator meet in person, they are coerced to engage in commercial sex acts, with the trafficker threatening to expose the images if the child refuses to comply.

Using artificial intelligence, fake images or videos can be created from content posted on social media sites and used to extort victims.

In cases of labor trafficking, the trafficker will post a fake job listing on job boards to attract potential victims. The trafficker will ask the potential victims to post their application online and conduct an interview, often acquiring personal information such as social security numbers, later using them for identity theft. Often, they entice the victim to leave their home country for a job overseas.

Click here to learn more.

Selling Children on Social Media

Traffickers often use covert communication when advertising or selling their victims. This may include coded language, hidden messages, emojis, or encrypted messaging to conceal their activities. Encryption is a technology that protects information by converting it into unreadable code that cannot be deciphered easily by unauthorized people.

These posts carry hidden meanings understood only by those involved in trafficking. Traffickers often post publicly without arousing suspicion while facilitating the selling of children.

Traffickers may also share social media accounts or profiles with others in their trafficking network to sell children.

Traffickers also communicate and are paid covertly when they commit the crime of "virtual child sex trafficking," which occurs when an offender in the United States sends a digital payment to a trafficker in another country. The trafficker will sexually abuse a child in front of a web camera while the offender in the United States watches a livestream of the abuse.

Traffickers will use live video streaming services and platforms to broadcast sexual abuse or exploitation of child victims in real-time. These streams are available to viewers who pay for access or belong to a closed group or forum via social media with the traffickers. Within these closed spaces online, traffickers can also share information about potential victims, exchange contact details, negotiate deals, and watch online child sexual abuse at a distance. Child sexual abuse, depicting the rape of infants and toddlers, bondage, humiliation through sexual assault, including self-mutilation, youth-on youth abuse, and child-on-child abuse, as well as bestiality, are not uncommon.

Finally, online child sex offenders are increasingly moving to the Dark Web. The Dark Web is a series of anonymous networks that prevent the use of traditional means to detect, investigate, and

Younger and younger children are being targeted "on an industrial scale" by internet groomers. In 2021, there was a three-fold increase in imagery showing 7–10 year olds targeted and groomed by internet predators.

(https://enough.org/stats_exploitation)

prosecute online child sexual exploitation offenses. The sites often expand rapidly. One site obtained 200,000 new members within its first four weeks of operation. Some sites require users to pay a fee to gain access, generally using cryptocurrency payments, commercializing the abuse suffered by victims whose images are trafficked. Others require new or prospective members to provide newly produced Child Sexual Abuse Material, pushing offenders further into their abuse of children. Though these sites sit within the Dark Web, they are readily accessible to anyone.

Click here to learn more.

Children's Brains and Social Media

Children are especially vulnerable to grooming by traffickers online. Children's brains undergo a fundamental shift starting around ten years old, which compels them to seek social rewards, especially attention and peer approval. Receptors for the "happy hormones" oxytocin and dopamine multiply in a part of the preteen brain, which makes them extra sensitive to attention and admiration from others. A dopamine and oxytocin rush occurs whenever the child experiences a "social reward" such as attention or a "like" from a friend. Therefore, classmates, friends, or people they have never met can deliver or withhold rewards in the form of "likes," views, and follows. The public nature of the internet also plays a role in the experience of "social reward."

This makes any child posting online vulnerable to bullying, exploitation, grooming, and trafficking.

Social media platforms like Instagram, YouTube, TikTok, and Snapchat have been increasingly linked to mental health problems, including anxiety, depressive symptoms, and body image concerns.

Adults may also experience this sense of reward when on social media. Still, they usually have a sense of selfhood that relies less on feedback from others, and they can usually better regulate their emotional responses.

*Traffickers will create accounts impersonating their victims and often spread lies or rumors about them on social media, referred to as "outing." Traffickers will also use social media to stalk victims, even when they can escape the life. Click **here** to learn more.*

Signs of Human Trafficking in Youth

Trafficking can be hard to recognize, especially for the people being groomed or trafficked. You can help keep children and teens safe by paying attention to changes in their behavior, such as:

- Avoiding you, friends, and family and having new friends they met online or older friends
- Frequent tardiness or absence from school or work- some children and teens are victimized after school or when they should be in school
- Sleeping often when they're home and staying out late or all weekend—many child victims of sex trafficking still live at home
- Loss of interest in things they once enjoyed
- Having or bragging about money, expensive items, or traveling with no reasonable explanation
- Starting to use drugs or alcohol
- Frequent injuries with no reasonable explanation
- Having secret online accounts- Children sometimes have two accounts on a social media platform. One, they show their parents if they check their phones; one, they do not.

These may be signs of human trafficking or peer bullying, or they could struggle with mental health and difficulties at home or school. This list is not inclusive.

Advocacy

Social Media Safety for Kids

Survive and Thrive Advocacy Center recommends that the most effective way to keep your kids safe online is to become more involved with the youth's online world. Warn them that someone might pretend to be a friend and be a trafficker, how traffickers target their victims, and red flags to watch out for.

Encourage them to speak with you if they feel uncomfortable or unsure about an online interaction or if anyone asks them for personal information.

Educate them on the dangers of "oversharing" online. Remind them that social media is not a personal photo album, and that people online are not always who they say they are.

Warn the child to refrain from posting information such as their full name, date of birth, contact number, or address.

Also, warn the child never to post pictures they would not want the world to see and never share the places or times of where you will be going nor add friends they do not know. Using Artificial Intelligence, traffickers can post fake images or videos from content posted on the child's social media sites.

Click here to learn more.

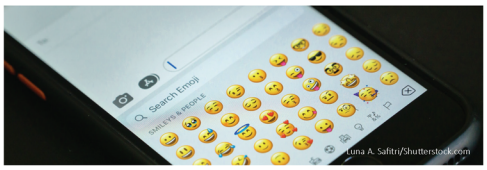


The Hidden Language

Children and teens use emojis as complete sentences; many have meanings unknown to many adults. The website Gabb recommends always considering the context when viewing an emoji.

Does a victory sign mean a team won a soccer game, or is it intended to be a vagina? A bowl of noodles could be a request for a naked photo, and broccoli sometimes means marijuana. Emojis of food, expressions, and parts of the body form the foundation of explicit and drug-related language.

Recognizing emojis' possible meanings is necessary to help keep children away from potential perpetrators. Please click here to access an Emoji meaning chart.



Recommendations for Social Media Platforms

Social media platforms often have policies and mechanisms in place to combat trafficking activities, which include collaborating with law enforcement to identify and apprehend offenders involved in the sex trafficking of children.

All social media platforms use algorithms, which are designed to determine the type of content that users see on their feeds or timelines. The algorithms collect data about users' behaviors, actions, and preferences based on the accounts they follow, the content they engage with, their search history, and demographic details, and these can be used to detect human trafficking and sex trafficking of children while ensuring the right to privacy.

Social media platforms can also develop proactive detection algorithms. These algorithms can analyze content and user behavior to identify patterns suggestive of child sexual exploitation.

Additionally, platforms should establish easily accessible reporting mechanisms that enable users to report content or activities that are suspected of child trafficking or exploitation. Moreover, anonymous reporting options encourage users to report suspicious activities without reluctance or fear.

Click here to learn more.

Livestreaming Child Sexual Exploitation (LCSE)

Livestreaming on social media has become extremely popular among children and adolescents. Livestreaming allows a user to produce real-time video broadcasted over an online social media platform, whether viewed publicly or by a restricted audience. Live.me and Omegle are among the most popular platforms for livestreaming, along with livestreaming capabilities on mainstream platforms like Facebook, Instagram, TikTok, YouTube, Snapchat, and Twitch. Most platforms have a chat feature where users can interact with viewers of their content.

Livestreaming child sexual exploitation (LCSE) occurs when a perpetrator coerces a child victim to engage in sexually explicit conduct in real-time to one or more viewers.

There are generally three types of LCSE: child "self-generated," offender-streaming, and virtual child sex trafficking. Child "self-generated" exploitation occurs when the trafficker coerces a child to engage in sexually explicit conduct on a live stream, usually from the child's bedroom or a bathroom.

Offender-streaming exploitation occurs when a perpetrator, usually a family member or friend, sexually abuses a child in person while live-streaming the abuse to viewers. The viewers typically do not know each other, and they often participate in the activity by requesting that specific sex acts be committed.

Finally, virtual child sex trafficking is when offenders pay to watch while another offender sexually abuses a child in person or offenders pay a victim directly to create "self-generated" Child Sexual Abuse Material (CSAM). Because of the interactive nature of live-streaming platforms, offenders can request specific sexual abuse acts for an additional cost. Payment is usually made digitally. This offense often involves offenders in the United States and facilitators and children in foreign countries. Children may be transported from rural areas of that foreign country to urban settings to fulfill the demand.

The trauma caused by LCSE and other forms of online child exploitation is compounded by the victim's knowledge that documentation of their abuse will live on the internet in perpetuity. Victims of online child sexual abuse must receive robust, ongoing victim services to aid in their healing.

IJM Summary Report Online Sexual Exploitation of Children

Over the past three months, law enforcement was not able to investigate nearly 100,000 IP addresses that are known to be downloading and sharing CSAM due to limited resources.

Senate Judiciary Hearing, Feb. 14, 2023



CD9 – Code 9: Parents in Room
KPC – Keeping Parents Clueless

Parents can go to several websites to decipher what their teen is saying on their cell phone text messages or social media sites. Please click here to access Text Lingo: Secret Language of Teens *for an up to date list of their coded language.*

Advocacy

Image-Based Sexual Abuse

Image-based sexual abuse (IBSA) is a criminal offense. Image-based sexual abuse includes the creation, theft, extortion, threatened or actual distribution, or any use of sexualized or sexually explicit materials without the consent of the person depicted.

Sexting is the sharing and receiving of sexually explicit messages and nude or partially nude images via cell phone, which an estimated 40 percent of teens engage in. Sexts may be sent as regular text messages through apps like Snapchat and WhatsApp or online games. Sexting, when involving minors, is legally classified as child sexual abuse, which is a serious federal crime in the United States. It is illegal to produce, possess, or distribute any visual depiction of sexually explicit conduct involving a minor.

Teens may "sext" for a variety of reasons, and some may feel pressured into sexting by online or offline boyfriends or girlfriends who may threaten to break up with them if they don't send a picture. They often rationalize that sending photos to one person won't hurt.

Unfortunately, many teens discover that someone they sent a sext to has forwarded that image to others without their knowledge. The consequences can be academically, socially, and emotionally devastating. Sexting usually violates school policy. The image may be seen by college admissions personnel or by potential employers. In some cases, the teen is charged for sending nude photos.

Image-based sexual abuse, sometimes referred to as 'revenge porn,' may be shared on specialized 'revenge porn' websites, on social media, via email, text, or messaging services, or shared with specific individuals, such as the victim's family, classmates, or employers. The images may also be shared offline. Sharing pictures or videos that have been photoshopped or otherwise altered in any way may also be considered image-based sexual abuse.

Consent is required at two stages: when the image or video is taken and again when it is shared with any third party. In the case of sexting, the person freely shares the pictures of themselves but is usually unaware that the images have then been shared or may not give consent to the photos being shared. Even in cases of self-generated explicit materials, it is never the victim's fault when their trust is broken and they are abused.

In some cases, the first level of consent is missing. For instance, the victim may have been unaware the image was taken, coerced into sharing the image, or the perpetrator may have stolen or hacked the image.

At times, the images are used to blackmail those depicted or to coerce them to send more sexually explicit photos. The perpetrator may threaten to harm your friends or relatives by using the images or other information they have obtained from you unless you comply with their demands. This is referred to as sextortion. Globally, sextortion has become a serious threat to young people. Adolescents are more susceptible to sextortion because of their developmental stage. They take more risks, struggle to control their impulses and desires, and are more easily swayed by peer pressure.

Child sextortion is becoming increasingly common—as far back as 2016, the U.S. Justice Department identified it as "the most important and fastest-growing cyberthreat to children." According to Enough is Enough, ninety percent of sextortion victims are teenage boys. Please click here to view their webinar on *Sex Trafficking in the United States: What every parent needs to know*. Boys are significantly less likely to tell their parents they are victims of sextortion than girls. Moreover, LGBTQ+ teens are more than twice as likely to be victims of sextortion as their heterosexual and gender-conforming peers.

Moreover, a "fake hacker" may claim that they have hacked into your child's device and found inappropriate sexual images. They will demand money or more images.

Other forms of image-based sexual abuse may include the non-consensual use of a person's images for the creation of photoshopped/artificial pornography or sexualized materials intended to portray a person, referred to as "cheap fake" or "deepfake" pornography. Also, IBSA includes the non-consensual recording of images or videos, including so-called "down blousing," "upskirting," or secret recordings in places such as restrooms and dressing rooms.

No one is free from the threat of IBSA.

Solutions must include stronger legislation, avenues for victims to sue in civil litigation, online platform responsibility to remove non-consensual explicit imagery swiftly, and survivor-centered removal forms. The laws regarding image-based sexual abuse vary from country to country.

For more information, please click here.

Action

How can YOU impact policy change?

The National Center on Sexual Exploitation (NCOSE) publishes a "Dirty Dozen List" each year, a campaign exposing twelve mainstream entities that enable and even promote and profit from sexual abuse and exploitation.

In 2023, all twelve were tech platforms involved in some way in facilitating the sexual exploitation of children and adults. NCOSE offers numerous opportunities for action anyone can take to advocate for and impact policy change with these platforms.

In past years, this campaign has yielded major victories at Google, Netflix, TikTok, Verizon, and many more.

Please find a brief description below of social media platforms and action alerts to which you can respond.

- Please click here to express concern with the rampant number of adults grooming minors for sexual abuse online and Discord's lack of robust safety features that allow such abuse to not only happen regularly but actively thrive in Discord's environment.
- Please click here to insist Instagram prioritizes child safety!
- Please click here to tell Reddit to delete all images of sexual violence from its platform.
- With Apple's nearly limitless resources, there can be no excuse for the deception of consumers and caregivers on such a massive scale: age ratings and descriptions mislead parents about the content, risks, and dangers to children on available apps. Please click here to urge Apple to fix its app age ratings.
- Tell Snapchat to step up child safety measure by clicking here.

Click here to learn more.

***Bark** scans your child's text messages, 30+ social media apps, web browsers, emails, and other online activity, keeping you informed, and your child protected. Bark can also help you manage screen times, block websites and apps with inappropriate content, and track location.*

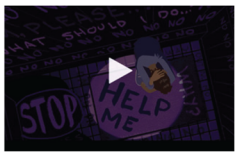
Danger Warning! Social Media and Sex Trafficking Recruitment

Social media is increasingly being exploited to contact, recruit, and sell children and youth for sex. Some traffickers use online ads to target victims, or they will send friend/follow requests to young people in their region, proceed to strike up a conversation and develop online friendships. These conversations may start friendly and innocent, but then the traffickers will begin to use manipulation to charm their potential victims. The online friendship quickly evolves into a romantic relationship, and this is where the sextortion techniques are used to lure their victims into human trafficking.

Please click here to view this webinar sponsored by Survive and Thrive in affiliation with Big Bend Coalition Against Human Trafficking and the International Rescue Committee.

With the fall 2023 iOS update, Apple added a feature to automatically blur images and videos containing nudity for children 12 and under in iMessage, FaceTime, AirDrop, and Photos Picker. This tool is also available for teens and adults as an opt-in feature. Previously, this blurring feature had to be turned on by parents, was unavailable to anyone over 13, only detected still images, and only worked in iMessage.

Action



Please click here to view a short video by the Blue Campaign of "Mia" meeting her trafficker over social media and being groomed.

NATIONAL CENTER on SEXUAL EXPLOITATION

Please click here to view video resources by the National Center on Sexual Exploitation on the impact of social media on child abuse, pornography, sexual violence and human trafficking.



When the Surgeon General issued an advisory on smoking in 1964, it became a priority to regulate Big Tobacco companies for the sake of public health. Now, we're facing a new crisis: social media's impact on the mental health and safety of kids. It must now be a priority to regulate Big Tech to protect our children's mental and emotional health - even their very lives. Please click here to view this 30 second PSA.

Sign the Petition to Show You Support John Doe

When John Doe was 13 years old, he was exploited by sex traffickers into creating sexually explicit images, which were later posted on Twitter. Twitter refused to remove or block the content depicting the sexual exploitation of John Doe, who was clearly and demonstratively a minor and continued to profit from its distribution knowingly.

It is on behalf of John Doe and countless other survivors like him that the National Center on Sexual Exploitation Law Center, along with The Haba Law Firm and The Matiasic Law Firm, has brought a lawsuit against Twitter.

Please click here to sign the petition in support of John Doe and all children exploited online.

Steps to Protect Children

Survive and Thrive recommends the following measures to help protect your children. Check the kids' devices frequently and thoroughly, including activity, messages, and contacts. Set appropriate parental control, such as age restrictions for downloading apps and time restrictions. At times, the child may have two separate accounts on a social media site. Be sure to check for the second site.

Educate yourself on how cybercrimes occur against children for sexual purposes. Once the online methods and tactics traffickers use are known, it is easier to detect signs of online exploitation.

Most importantly, maintain a trustful relationship with your children. While safety features are helpful, you should rely more on your relationship with them than filters and other safeguards. Also, a child's safety is more important than their privacy. This is not a trust issue; it is a desire to love and protect your child.

NATIONAL CENTER FOR MISSING & EXPLOITED CHILDREN®

24-Hour Call Center:

To report information about a missing or exploited child call our 24-Hour Call Center:

1-800-THE-LOST (1-800-843-5678)

Report child sexual exploitation online at CyberTipline.org.

ALLIANCE TO END HUMAN TRAFFICKING

Founded and Supported by U.S. Catholic Sisters

Click on the links below to visit the websites of our sponsors

- Adorers of the Blood of Christ
- Adrian Dominicans
- Benedictine Sisters of Chicago
- Benedictine Sisters of Mount St. Scholastica, Atchison, KS
- Benet Hill Monastery
- Congregation of Notre Dame
- Congregation of Sisters of St. Agnes
- Congregation of S. Joseph
- Daughters of Charity, Province of the West
- Daughters of Charity, Province of St. Louise
- Daughters of the Holy Spirit
- Dominican Sisters of Houston, TX
- Dominican Sisters of Mission San Jose, CA
- Dominican Sisters of Peace
- Dominican Sisters of San Rafael, CA
- Dominican Sisters of Sinsinawa, WI
- Dominican Sisters of Springfield, IL
- Felician Sisters of North America
- Franciscan Sisters of Peace
- Franciscan Sisters of Perpetual Adoration
- Franciscan Sisters of the Sacred Heart
- Holy Spirit Missionary Sisters
- Institute of the Blessed Virgin Mary
- Marianites of Holy Cross
- Maryknoll Sisters
- Medical Mission Sisters
- Medical Missionaries of Mary
- Missionary Sisters of the Society of Mary
- Northern California Catholic Sisters Against Human Trafficking
- Our Lady of Victory Missionary Sisters
- Presentation Sisters, Aberdeen
- Presentation Sisters, San Francisco
- Racine Dominicans
- Religious of the Sacred Heart of Mary
- Religious Sisters of Charity
- School Sisters of Notre Dame, North America
- School Sisters of St. Francis of Christ the King
- Sisters of Bon Secours
- Sisters of Charity of Cincinnati
- Sisters of Charity of Halifax
- Sisters of Charity of Leavenworth
- Sisters of Charity of New York
- Sisters of Charity of St. Joan Antida
- Sisters of Charity of the Blessed Virgin Mary
- Sisters of Charity of the Incarnate Word – Houston
- Sisters of Charity of Nazareth
- Sisters of Charity of Seton Hill
- Sisters of Christian Charity Mendham, NJ & Wilmette, IL
- Sisters of Mercy Catherine's Residence
- Sisters of Mercy of the Americas
- Sisters of Notre Dame of the United States
- Sisters of Notre Dame de Namur, USA
- Sisters of Providence, Mother Joseph Province
- Sisters of St. Dominic – Racine, WI
- Sisters of St. Francis of Clinton
- Sisters of St. Francis of Colorado Springs
- Sisters of St. Francis of Dubuque
- Sisters of St. Francis of Philadelphia
- Sisters of St. Francis of Redwood City
- Sisters of St. Francis of the Providence of God
- Sisters of St. Francis Rochester, MN
- Sisters of St. Joseph of Baden
- Sisters of St. Joseph of Carondelet
- Sisters of St. Joseph of Chestnut Hill Philadelphia
- Sisters of St. Joseph of Cluny, USA & Canada Provinces
- Sisters of St. Joseph of Concordia, KS
- Sisters of St. Joseph of Orange
- Sisters of the Blessed Sacrament
- Sisters of the Divine Savior
- Sisters of the Good Shepherd
- Sisters of the Holy Cross
- Sisters of the Holy Family
- Sisters of the Holy Names of Jesus and Mary
- Sisters of the Humility of Mary
- Sisters of the Precious Blood
- Sisters of the Presentation of the Blessed Virgin Mary
- Sisters of the Sacred Hearts
- Society of the Divine Savior
- Society of the Holy Child Jesus
- Society of the Sacred Heart
- Southern CA Partners for Global Justice
- St. Mary's Institute of O'Fallon
- Tri-State Coalition Against Human Trafficking & Slavery
- U.S. Ursuline Sisters of the Roman Union

The Anti-Trafficking Newsletter is dedicated exclusively to fostering an exchange of information among Alliance members, organizations and concerned persons collaborating to eliminate all forms of human trafficking. Click here to access previous issues of Stop Trafficking! To contribute information, please contact: stoptrafficking@feliciansisters.org. Editor: Maryann Agnes Mueller, CSSF. Layout & Design: Mary Francis Lewandowski, CSSF. Translated into Spanish by Amelia Breton, SBS.

Center for AI and Digital Policy

January 31, 2024

Chairman Dick Durbin
Ranking Member Lindsay Graham
U.S. Senate Committee on the Judiciary
G50 Dirksen Senate Office Building, Washington D.C. 20515

Re: CAIDP Statement for the Record: "Big Tech and the Online Child Sexual Exploitation Crisis"

Dear Chairman Durbin, Ranking Member Graham, and Members of the Committee,

We write to you regarding the hearing on "***Big Tech and the Online Child Sexual Exploitation Crisis.***"[1] We thank you for your continued leadership in protecting the American people from the risks of technology, holding Big Tech accountable particularly regarding online child safety. We write to highlight the serious risks to children further exacerbated through the unprecedented deployment of AI systems in children facing apps, online mediums, and social media platforms and urge you to take our recommendations into consideration.

The Center for AI and Digital Policy (CAIDP) is an independent research and education non-profit based in Washington, DC.[2] Our global network of AI policy experts and advocates advise national governments, international organizations, and Congressional committees regarding artificial intelligence and digital policy. Our President, Merve Hickok testified at the first congressional hearing on AI last year–"*Advances in AI: Are We Ready For a Tech Revolution*?"[3].

CAIDP routinely provides advice to Congressional Committees on matters involving AI policy[4], and we also submitted our statements to this Committee on "Oversight of A.I.: Rules for Artificial Intelligence"[5] and "Oversight of A.I.: Legislating on Artificial Intelligence".[6] We also publish the annual *Artificial Intelligence and Democratic Values Report*,[7] providing a comprehensive review of AI policies and practices in 75 countries.

---

[1] U.S. Senate Committee on the Judiciary, *Big Tech and the Online Child Sexual Exploitation Crisis,* (January 31, 2024), https://www.judiciary.senate.gov/committee-activity/hearings/big-tech-and-the-online-child-sexual-exploitation-crisis.
[2] CAIDP, *About*, https://www.caidp.org/about-2/.
[3] Testimony and statement for the record of CAIDP President Merve Hickok, *Advances in AI: Are We Ready For a Tech Revolution?*, House Committee on Oversight and Accountability, Subcommittee on Cybersecurity, Information Technology, and Government Innovation (March 8, 2023), https://oversight.house.gov/wp-content/uploads/2023/03/Merve-Hickok_testimony_March-8th-2023.pdf.
[4] CAIDP, *Statements, https://www.caidp.org/statements/*
[5] CAIDP, *Statement to Senate Judiciary Committee on Oversight of AI,* (May 11, 2023), https://www.caidp.org/app/download/8457859963/CAIDP-Statement-SJC-AI-05112023.pdf
[6] CAIDP, *Statement to Senate Judiciary Committee on U.S. AI Act,* (September 12, 2023), https://www.caidp.org/app/download/8475524463/CAIDP-SJC-09112023.pdf
[7] CAIDP, *Artificial Intelligence and Democratic Values* (2023), https://www.caidp.org/reports/aidv-2022/.



In brief, our recommendations to this Committee are:

1) Address the grave risks of generative AI in social media

2) Enact legislation establishing accountability measures for AI systems including mandating synthetic content labelling, and for protecting minors from AI-enabled sexual abuse, extortion, and pornography.

Recommendation 1. Address the risks of Generative AI in social media

The deployment of generative AI tools particularly in apps and social media, especially those targeted at children, is an alarming development and must be reined in by Congress. Generative AI refers to the ability for machines to generate new, incredibly realistic, and detailed text, code, images, audio, and more. The commercialization of generative AI tools and consumer facing applications has far outpaced that of social media.

It is well established by researchers, advocates, and government that social media companies thrive off hooking minors on their products.[8] A recent study by the Harvard T.H. Chan School of Public Health estimated that ads targeted at American children generated nearly $11 billion in revenue in 2022 for just Meta, Snapchat, TikTok, X, and YouTube.[9] As such, these companies have aggressively integrated various generative AI capabilities into their products, even though general-purpose AI chatbots like ChatGPT and Bard either require parental approval for teenagers or exclude them outright.[10]

For instance, Snapchat includes an OpenAI-powered chatbot assistant in its app–*My AI*, which is specifically designed to grip the app's primarily minor users. The chatbot is pinned to the top of a user's messages page and cannot be removed on the app's free version.[11] When a user messages My AI, the app-generated-bitmoji for the AI is shown as thinking and writing a response like a human messenger would. The chatbot routinely includes emojis in its text responses, uncharacteristic of most AI chatbots, but common among young people. These design choices seek to anthropomorphize the

[8] Melissa Henson, *Big Tech is preying on children for profit, and Congress needs to stop it*, The Hill (January 18, 2024), https://thehill.com/opinion/technology/4411420-big-tech-is-preying-on-children-for-profit-and-congress-needs-to-stop-it/.
[9] Harvard T.H. Chan School of Public Health, *Social media platforms generate billions in annual ad revenue for U.S. youth,* (December 27, 2023), https://www.hsph.harvard.edu/news/press-releases/social-media-platforms-generate-billions-in-annual-ad-revenue-from-u-s-youth/.
[10] Jeffrey R. Young, *Teens Need Parent Permission to Use ChatGPT. Could That Slow Its Use in Schools?*, EdSurge (November 2, 2023), https://www.edsurge.com/news/2023-11-02-teens-need-parent-permission-to-use-chatgpt-could-that-slow-its-use-in-schools.
[11] *How do I unpin or remove My AI with Snapchat+?*, Snap Inc., https://help.snapchat.com/hc/en-us/articles/13387249333780-How-do-I-unpin-or-remove-My-AI-with-Snapchat. Accessed January 30, 2024.

algorithm and deceive children,[12] many of whom will not understand what an AI is, AI's risks and harms, or how their conversations with the chatbot are used by Snap. Snap disclaims any liability in the following terms "My AI may include biased, incorrect, harmful or misleading content"[13]; and suggests that users should independently verify any advice it gives before acting on it.[14]

Indeed, this use can be, and already is, quite dangerous:

- ***Promoting dangerous behavior.*** As highlighted in an open letter from Senator Bennet to tech CEOs, Snapchat's My AI was used to instruct a child on how to cover up a bruise ahead of a visit from Child Protective Services and provided suggestions to a 13-year-old girl on how to lie to her parents about an upcoming trip with a 31-year-old man.[15]

- ***Amplification of hallucinations, bias, and extremism.*** These well-documented AI risks[16] are especially problematic for young people. Image-generation tools like Stable Diffusion consistently generate white, male-centric images that could reinforce stereotypes and bias among children.[17] An unwitting child on WhatsApp could innocently prompt its image generation software with "Palestinian boy" and see violent, graphical images of children holding guns.[18]
  Other models like OpenAI's GPT-4V doesn't understand hate symbols like "the modern meaning of the Templar Cross (white supremacy) in the U.S." despite best-attempt safety training.[19] We must protect children from these harms, and this necessitates additional regulation of AI systems, and enforcement by the FTC, in certain contexts.

---

[12] Sharon Goldman, *Sen. Murphy's tweets on ChatGPT spark backlash from former White House AI policy advisor,* VentureBeat (March 28, 2023), https://venturebeat.com/ai/sen-murphys-tweets-on-chatgpt-spark-backlash-from-former-white-house-ai-policy-advisor/.
[13] Bernard Marr, *Snapchat Debuts ChatGPT – Powered Snap AI: But is it safe for kids?*, Forbes (April. 26, 2023), https://www.forbes.com/sites/bernardmarr/2023/04/26/snapchat-debuts-chatgpt- powered-snap-ai--but-is-it-safe-for-kids
[14] CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (July 10, 2023), pg. 26, https://www.caidp.org/cases/openai/
[15] *Following Early Reports of Potential Harmful Content from AI Chatbots, Bennet Urges Tech CEOs to Prioritize Young Americans' Safety*, Michael Bennet, U.S. Senator for Colorado (March 21, 2023), https://www.bennet.senate.gov/public/index.cfm/2023/3/bennet-calls-on-tech-companies-to-protect-kids-as-they-deploy-ai-chatbots.
[16] FTC, *Consumers Are Voicing Concerns About AI,* (Oct. 3, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/10/consumers-are-voicing- concerns-about-ai
[17] Leonardo Nicoletti and Dina Bass, *Generative AI Takes Stereotypes and Bias From Bad to Worse*, Bloomberg (2023), https://www.bloomberg.com/graphics/2023-generative-ai-bias/.
[18] Johana Bhuiyan, *WhatsApp's AI shows gun-wielding children when prompted with 'Palestine'*, The Guardian (November 3, 2023), https://www.theguardian.com/technology/2023/nov/02/whatsapps-ai-palestine-kids-gun-gaza-bias-israel.
[19] OpenAI, *GPT-4V(ision) System Card*, https://cdn.openai.com/papers/GPTV_System_Card.pdf. *See also,* CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (November 14, 2023), pg. 16, https://www.caidp.org/cases/openai/



- ***Privacy Violations.*** AI will also exacerbate the already substantial privacy concerns regarding children's use of social media. The intrusive nature of chatbots and social media will invariably extract personal and sensitive information – from text conversations, images, and other information – that will then be used to train models and target advertisements and content at our children.[20] A report by the *Scientific American* shows that "OpenAI fine-tunes its models based on user interactions with its chatbots."[21]

- ***Mental Health Risks.*** Chatbots and other generative AI systems present a natural attraction for children to use social media products more. However, the current use of social media is already a disaster for our children. Social media is a key driver of a mental health crisis among American youth,[22] resulting in 40% of teens feeling persistently sad or hopeless in 2021 and the American Academy of Pediatrics declaring a national emergency for child and adolescent mental health.[23]

We highlighted these and other risks in our complaint to the FTC against OpenAI:[24]

> "ChatGPT has demonstrated negative effects of providing misinformation, incorrect advice, creation of bullying content. The inclusion of a conversational feature in social media apps used predominantly by children is likely to result in higher amounts of data scraped from children, increased screen-time and endanger online child safety." [25]

The FTC issued a civil investigative demand on OpenAI in July of last year but almost 1 year since the filing of the complaint, the FTC is yet to issue any meaningful order. History shows that the longer the FTC delays, the more difficult it is to establish the necessary guardrails. Inaction by the agency is costly.

---

[20] AI Review Team, *My AI*, Common Sense Media (October 19, 2023), https://www.commonsensemedia.org/ai-ratings/my-ai.
[21] Lauren Leffer, *Your Personal Information Is Probably Being Used to Train Generative AI Models, Scientific American,* (October 19, 2023), https://www.scientificamerican.com/article/your- personal-information-is-probably-being-used-to-train-generative-ai-models/ ; CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (November 14, 2023), https://www.caidp.org/cases/openai/
[22] Chris Griswold, *Big Tech Is Exploiting Kids Online. Congress Has To Step In*, Newsweek (November 6, 2023), https://www.newsweek.com/big-tech-exploiting-kids-online-congress-has-step-opinion-1840276.
[23] Moriah Balingit, *'A cry for help': CDC warns of a steep decline in teen mental health*, The Washington Post (March 31, 2022), https://www.washingtonpost.com/education/2022/03/31/student-mental-health-decline-cdc/.
[24] CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (July 10, 2023), https://www.caidp.org/cases/openai/
[25] *Id,* at pg. 25, 26. *See also,* Open AI, *The GPT-4 System Card* (March 15, 2023), https://cdn.openai.com/papers/gpt-4-system-card.pdf; The Lancet, *ChatGPT: Friend or Foe,* Editorial (February 6, 2023), https://www.thelancet.com/action/showPdf



To address the risks of generative AI in social media, we urge this Committee to exercise its oversight authority and call upon the FTC to conclude its investigation into OpenAI. We also urge this Committee to exercise oversight on the FTC's enforcement plans in relation to consumer harms, arising out of the harmful, exploitative, and deceptive commercial practices of AI companies developing generative AI tools, and data partnerships with social media companies deploying them downstream.

Recommendation 2. Enact legislation establishing accountability measures for AI systems and prohibiting Sexual Deepfakes and Violence

Generative AI now has the capability to create extremely realistic image and video depictions and synthetic voice content of specific people, colloquially known as "deepfakes."[26] For this hearing, we specifically focus on the extreme risks posed by deepfakes to children and sexual exploitation.

Deepfakes are already used to construct nonconsensual pornographic content at alarming scale and ease,[27] and many are easily available online.[28] But the technology has no age restrictions. Indeed, many children – most, girls – are already falling victim to this vile process.[29]

The Internet Watch Foundation has already found nonconsensual deepfake sexual abuse images of famous children and new images of children who were abused in the past.[30] They also found that abusers were "sharing tips and marveling about how easy it was to turn their home computers into factories for generating sexually explicit images of children of all ages."[31]

However, children are also becoming perpetrators of these harms. Just a few months ago, teenage boys in New Jersey generated nonconsensual pornographic photos of their classmates using

[26] Daniel I. Weiner & Lawrence Norden, *Regulating AI Deepfakes and Synthetic Media in the Political Arena*, Brennan Center for Justice (December 5, 2023), https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena.
[27] Angela Chen, *Forget fake news—nearly all deepfakes are being made for porn*, MIT Technology Review (October 7, 2019), https://www.technologyreview.com/2019/10/07/132735/deepfake-porn-deeptrace-legislation-california-election-disinformation/.
[28] Kat Tenbarge, *Google and Bing put nonconsensual deepfake porn at the top of some search results,* NBC News (January 11, 2024), https://www.nbcnews.com/tech/internet/google-bing-deepfake-porn-image-celebrity-rcna130445.
[29] Matt O'Brien and Haleluya Hadero, *AI-generated child sexual abuse images could flood the internet. Now there are calls for action,* The Associated Press (October 25, 2023), https://apnews.com/article/ai-artificial-intelligence-child-sexual-abuse-c8f17de56d41f05f55286eb6177138d2.
[30] *Id.*
[31] *Id.*



AI.[32] But the victims had scant means for recourse, as "the incident would likely amount to nothing more than a 'cyber-type harassment claim,' rather than a more serious crime like child pornography."[33]

Social media can spread deepfakes far and wide, as recently occurred with singer-songwriter Taylor Swift,[34] furthering the harm for young people. Deepfakes have also been used to further "sextortion," blackmailing victims by threatening to share the manufactured images.[35] Sextortion guides are widely available on social media like TikTok, Instagram, Snapchat, and YouTube and are only becoming more prevalent.[36]

Congress must act swiftly and decisively on this issue. Although mainstream image generation products disallow sexually explicit uses, other software are widely available because the fundamental AI technology is open sourced.[37]

We encourage this Committee to further act on its mandate to enact legislation protecting minors from nonconsensual deepfake pornographic material. While there has been much discussion on the tension between first amendment rights and expanded criminalization of the creation, use, and dissemination of child sexual abuse material, the proliferation of AI systems raise unprecedented risks to online child safety that must be addressed through amending existing laws or enacting new laws.

We also urge this Committee to move forward with AI legislation. There have been several bills proposed in Congress, but none have moved forward. **We endorse the Blumenthal-Hawley Bi-Partisan Framework for a US AI Act.[38]**

---

[32] Tate Ryan-Mosley, *A high school's deepfake porn scandal is pushing US lawmakers into action,* MIT Technology Review (December 1, 2023), https://www.technologyreview.com/2023/12/01/1084164/deepfake-porn-scandal-pushing-us-lawmakers/.
[33] *Id.*
[34] *Taylor Swift deepfakes spread online, sparking outrage*, CBS News (January 26, 2024), https://www.cbsnews.com/news/taylor-swift-deepfakes-online-outrage-artificial-intelligence/.
[35] Jack Guy, *Outcry in Spain as artificial intelligence used to create fake naked images of underage girls,* CNN (September 20, 2023), https://www.cnn.com/2023/09/20/europe/spain-deepfake-images-investigation-scli-intl/index.html.
[36] Lora Kolodny, *Sextortion training materials found on TikTok, Instagram, Snapchat and YouTube, according to new report,* NBC News (January 27, 2024), https://www.nbcnews.com/tech/internet/sextortion-yahoo-boys-snapchat-tiktok-teen-wizz-rcna134200; David Finkelhor, Heather Turner, and Deirdre Colburn, *Prevalence of Online Sexual Offenses Against Children in the US,* JAMA Network Open (October 2022), https://doi.org/10.1001%2Fjamanetworkopen.2022.34471.
[37] O'Brien and Hadero, *supra* note 17.
[38] Senator Richard Blumenthal & Senator Josh Hawley, *Bipartisan Framework for U.S. AI Act*, https://www.blumenthal.senate.gov/imo/media/doc/09072023bipartisanaiframework.pdf

Center for AI and Digital Policy

However, to address specifically, the risks to child safety, we recommend further refinements to the Blumenthal-Hawley framework and/or any other legislation this Committee would consider:

(i) a clear requirement to label all AI-generated images, audio, and video for ease of detection

(ii) private rights of action and/or criminal liability for creating, disseminating, or using nonconsensual deepfake pornography, image-based sexual abuse of minors

Thank you for your consideration of our views. We ask that this statement be included in the hearing record. We would be pleased to provide you and your staff with additional information.

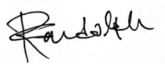
Sincerely yours,



Marc Rotenberg
CAIDP Executive Director



Merve Hickok
CAIDP President



Christabel Randolph
Law Fellow



Maanas Sharma
Research Assistant

CENTER FOR DEMOCRACY & TECHNOLOGY

January 30, 2024

The Honorable Dick Durbin
Chairman, Senate Judiciary Committee
U.S. Senate
Washington, D.C. 20510

The Honorable Lindsay Graham
Ranking Member, Senate Judiciary Committee
U.S. Senate
Washington, D.C. 20510

CC: Members of the Senate Judiciary Committee

Re: Senate Judiciary Committee's January 31, 2024, Hearing on "Big Tech and the Online Child Sexual Exploitation Crisis"

Dear Chairman Durbin, Ranking Member Graham, and Members of the Committee,

The undersigned organizations write to urge Committee members to focus their questions at tomorrow's hearing on child safety on ensuring the technology companies represented at the hearing are appropriately investing in identifying threats to young internet users and developing content-neutral tools to mitigate these threats including tools that empower young people to protect themselves and their privacy and to refrain from advancing legislation that, as currently drafted, undermines their rights.

We strongly agree with Committee members that ensuring young people are safe online is essential. The goal of ensuring safety is not and should not be at odds with supporting young people's autonomy to access information, use secure communication channels, and make decisions about their online experiences. Compromising the latter goals to achieve an illusory sense of safety will undermine young people's well-being as well as their fundamental rights. As many of our organizations have highlighted, some of the legislative proposals before the Senate bear this risk and threaten to jeopardize all internet users' access to information and privacy.

Access to online services is essential for young people's development
Young people rely on online services to stay in touch with friends and family, find and nurture interests, and develop their sense of self. Research has repeatedly shown that in today's world, social media and secure messaging services are integral for young people and their well-being.

A poll conducted by Pew Research Center in April 2023 found that the 80% of teens report that access to social media helps them feel more connected to their friends' lives, 71% of teens feel like social media is a place where they can be creative, and 67% of teens use social media to support them when they are going through tough times. In particular, marginalized youth such as LGBTQ+ teens and teenagers of color rely on online spaces to nurture their identity, find community, and exchange resources. LGBTQ+ teens use encrypted messaging services and disappearing messaging features offered by social media companies to confide in friends and trusted adults, especially when facing unsupportive or abusive situations at home.

Black-and-white assertions about the role social media plays in young people's lives risks obscuring the real ways young people use and rely on these services. The Surgeon General's advisory guidance on youth mental health says that social media offers both benefits and drawbacks for young people and that their benefits may be especially salient for marginalized youth. As youth advocate Sarah Philips wrote in Teen Vogue, "the most under-cited part of the [Surgeon General's] report is actually the most important… LGBTQ+ youth actually have better outcomes when they have access to social media and secure online communities."

Legislation being reviewed by this Committee could undermine young people's access to critical information and in turn, their well-being

Proposed legislation before this and other Committees could jeopardize young people's access to information and privacy, and potentially impact their well-being. Bills like the most recent public version of the Strengthening Transparency and Obligations to Protect Children Suffering from Abuse and Mistreatment Act (STOP CSAM Act), the Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2023 (EARN IT Act), and the version of the Kids Online Safety Act (KOSA) that passed through the Senate Commerce Committee in July 2023, contain strong incentives to use content filtering tools and/or eliminate end-to-end-encryption guarantees by weakening or introducing new liability standards for platforms that host user-generated content, some of it that is perceived to be harmful to children.

Creating or increasing the risk of liability that platforms face for hosting certain categories of content will inevitably lead reasonably risk-averse platforms to institute compliance measures, including filtering tools, that sweep far more broadly than those categories. Unclear liability standards such as a "duty of care" which requires companies to prevent a set of harms caused by the design of the platform may result in similar effects with companies developing and deploying aggressive content filters to reduce the visibility of lawful content that platforms fear enforcers may interpret as harmful. The use of these tools is likely to result in over-moderation of First Amendment-protected and developmentally-important speech, including about gender identity, sexual orientation, and reproductive health. The passage of SESTA-FOSTA is one

example where a broadly scoped carve-out of intermediary liability protections chilled speech, particularly by the LGBTQ+ community.

Websites focused on sex education and LGBTQ+ issues have been regularly swept up in attempts to block sexual content due to the imprecise nature of content filtering tools. A study conducted by Top10VPN in 2021 found that 92% of child-safety apps on Google Play were wrongly blocking at least one LGBTQ or sex-education website including the Trevor Project and It Gets Better as "adult content." An EFF analysis of a popular student monitoring tool, GoGuardian, found that a vast majority of sample websites were inaccurately flagged as potentially dangerous just for mentioning information about LGBTQ issues —even when those sites were merely educational or informative. A WIRED investigation in 2023 into commercial child-safety content filtering tools found that these tools also blocked access to essential information a young person may seek as part of their education and development including websites of Human Rights Watch and Amnesty International, artistic interpretations of The Hobbit, and even press coverage of school shootings. Widespread use of these tools to avoid the risk of liability resulting from legislation like the STOP CSAM Act will only lead to more censorship of essential and lawful speech affecting all internet users' fundamental rights.

Proposed legislation would have an outsized impact on all internet users, removing all users' access to secure communications channels

Some of the companies represented at the hearing offer secure end-to-end encrypted group messaging services, which according to a recent study by the Center for Democracy & Technology, young people aged 14 to 21 use to stay in touch with friends and family. Ensuring these group chats cannot be intercepted by malicious actors is paramount to the safety of young people, and all internet users. Yet, legislation, including STOP CSAM, EARN IT and KOSA, would incentivize providers not to offer encrypted services or to weaken encryption in order to facilitate additional surveillance of content. This surveillance will be done in the name of protecting child safety, but will instead undermine their right to privacy and that of all other internet users. Weakening or eliminating access to encrypted platforms would harm all internet users' right to private communications and in particular harm LGBTQ+ youth who depend on encrypted messages to confide in trusted adults and counselors, journalists who use encrypted messages to contact their sources, doctors who use it to speak with patients, domestic violence victims who rely on completely private communications to escape dangerous situations at home, and businesses discussing finances with clients.

Weakening encryption would also equip law enforcement with easier access to more information on our private communications and social ties, enabling increased government surveillance. For example, a woman in the UK received a 28-month sentence for taking abortion

pills based on her internet searches and private messages. All of us, including children, need to be able to communicate in the digital age without our conversations being spied on.

Young people want tools to empower themselves and to keep themselves safe
Young people have regularly called on technology companies to give them the ability to shape their own healthy, online experiences. What a healthy online experience looks like to one group of young people is likely to be different from another. When CDT researcher Michal Luria spoke with over 30+ young people as part of a study to understand the nature of unwanted encounters young people receive on direct messaging services, many told her that they relied on reporting, blocking, and privacy features to keep themselves safe. Many respondents asked for dynamic tools such as the ability to block certain words in their comments or receive notices when an individual outside of their network is trying to send a message. These tools, the participants aged 14 to 21 argued, helped them shape their online environment and keep themselves safe.

This autonomy is critical for young people's development according to child psychologists. In a separate paper published by the Journal of Pediatrics in 2023, researchers say that early independence amongst young people is critical for emotional development, future decision-making skills, and can even have better mental health outcomes. Policymakers can bolster this call from young people by asking the technology companies represented at the hearing about how their teams build user tools to promote teenage independence.

Passing legislation that imposes strong incentives for companies to create a one-size-fits-all approach to child safety and remove access to secure communication channels will do more harm than good for marginalized youth and all internet users.

If you have any questions, please do not hesitate to contact Aliya Bhatia, CDT (abhatia@cdt.org) or Jenna Leventoff, ACLU (jleventoff@aclu.org).

Sincerely,
Signatories
American Civil Liberties Union
Center for Democracy & Technology
Electronic Frontier Foundation
Fight for the Future



United States Senate
Committee on the Judiciary

WRITTEN TESTIMONY BY

End OSEAC Coalition's Survivors' Council,
End Online Sexual Exploitation and Abuse of Children Coalition

For a Hearing Entitled:
Big Tech and the Online Child Sexual Exploitation Crisis

Wednesday, January 31st, 2023

ENDOSEAC
COALITION TO PROTECT KIDS ONLINE

Dear Chairman Durbin, Ranking Member Graham, and esteemed members of the Senate Judiciary,

We, the members of the End Online Sexual Abuse and Exploitation of Children (OSEAC) Coalition Survivors' Council, write to fully endorse and demand support for radical change across social media platforms such as X, Meta, Snap, Discord, and TikTok. This change must occur immediately and through the support of legislation like the EARN IT Act, the REPORT Act, the SHIELD Act, and the STOP CSAM Act.

Social media and technology have become ever-changing realms. We have seen the growth of these respective platforms exponentially as change across social media occurs every second. As more people begin and continue to gain access to social media, it is without a doubt that this has created an incredibly harmful reality for children, parents, and society as a whole.

With the growth and accessibility of social media comes the ability for predators and other dangerous individuals to have easier access to children. This accessibility also causes childhood sexual abuse materials (CSAM) to be further disseminated across multi-layered platforms. According to the National Center for Missing and Exploited Children (NCMEC), over 85 million images and videos of child sexual abuse were found online, indefinitely remaining on the Internet.[1] Furthermore, according to the Federal Bureau of Investigation, 50% of the victims of online sexual exploitation are children ages 12 to 15. These are only two of the plethora of alarming statistics regarding CSAM and its existence online.

We can only imagine these numbers will grow as social media use develops. For this reason, the time to take action that mitigates these risks is now. For far too long, children have suffered horrible consequences due to a lack of accountability by major corporations and their blanket immunity. The plight of survivors is only heightened by mass apathy regarding the sexual assault, grooming, and exploitation of children online. Why is it that the safety of social media users, especially as the number of children accessing these platforms grows daily, is not prioritized? The protection of children against sexual predators must be of the utmost importance to Congress and the Social Media Companies that testified before you. We want to say there cannot be any more delays regarding change. We need to pass legislation that prioritizes online safety. The alarming number of children falling victim to exploitation demands government intervention, and we implore you to not prolong this process any further, as time is of the essence.

To truly see an end to Online Sexual Exploitation (OSEC), every sector has to play a role. It is our collective responsibility to protect the innocence and well-being of our children in the digital age. We call upon parents, social media companies, Congress, schools, and civil society to take immediate action to create a safer online environment for our children.

[1] https://www.missingkids.org/cybertiplinedata

END OSEAC
COALITION TO PROTECT KIDS ONLINE

To the Tech CEOs who testified, we demand that you invest in tools that provide practical online moderating abilities and safety by their design. This includes but is not limited to actively monitoring social media, committing to removing CSAM, and accurately reporting its existence on these platforms. Furthermore, it means investing in parental controls so a trusted adult can access their child's social media for further protection and monitoring. It is concerning that predators can operate freely on your platforms, perpetrating these heinous crimes while you seemingly take no action. By failing to act, you become complicit in these atrocities. We implore you to do what is right and take steps to protect our children. This is an opportunity to be a leader in the movement and serve as an example of effective change.

According to our suggested actions, these platforms must consider, invest in, and immediately utilize online safety features. A safer, more equitable social media presence is possible for users everywhere and we can ensure a safer space for them.

Please remember that we are speaking from our own experiences. We, the members of the End OSEAC Survivors' Council, continue to be directly impacted by this issue and its pervasiveness. The horrific and dark truth about the online sexual exploitation of children is that it's an ongoing crime. Our images and videos continue to be found online and are a constant reminder of the horrors we went through. They are just one click away from the hands of predators here in the U.S. and across the globe. We are tired of seeing the retraumatization of our fellow survivors.

Today and every day, we cry out for justice. We hope that lawmakers and tech owners will hear our voices and ensure the safety of children on their platforms. Together, let us work towards a future where our children can grow up without fear, where their innocence is preserved, and where online platforms prioritize the safety and well-being of every child. The time for action is now.

Thank You,

End OSEAC Coalition Survivors' Council

End OSEAC Coalition Survivors' Council



The End Online Sexual Exploitation and Abuse of Children (OSEAC) Survivors' Council is a diverse lived experience experts all working to end sexual abuse and exploitation of children online. The council is a vital part of the END OSEAC Coalition, a U.S. advocacy coalition that aims to improve U.S. government policies and programs to better prevent and address the online sexual exploitation and abuse of children and provide appropriate support to survivors. Visit https://endoseac.org/ to learn more.



United States Senate
Committee on the Judiciary

WRITTEN TESTIMONY BY

Christine Almadjian,
End Online Sexual Exploitation and Abuse of Children (OSEAC) Coalition

For a Hearing Entitled:
Big Tech and the Online Child Sexual Exploitation Crisis

Wednesday, January 31st, 2023

ENDOSEAC
COALITION TO PROTECT KIDS ONLINE

Dear Chairman Durbin, Ranking Member Graham, and esteemed members of the Senate Judiciary,

My name is Christine Almadjian. I'm a graduate student at Columbia University and a member of the End Online Sexual Exploitation and Abuse of Children Coalition's Survivors' Council. It is with great honor that I stand here before you to share my opinions and thoughts, not forgetting the absolute urgency this issue requires from all these companies.

I may be 22, but I distinctly remember life at 11. As middle schoolers, many of us were entering a new realm of interaction as access to smartphones grew exponentially. Alongside the group chats with friends and interactive game apps came the hustle and bustle of social media. Some apps, such as Snapchat, Twitter (X), and Meta, were among the most popular. I remember my friends and I sitting around one another at lunch, retweeting posts about One Direction, and taking selfies with filters to send one another on Snapchat.

What we initially considered fun took very dark turns for many of us. As children, we did not have the foresight to detect dangerous situations the way we might now or the way I at least do. As we were invited to chat groups and private messaging avenues on the apps, exposure to predators became concurrent with our social media experience.

I can personally recall the predators that forced their way into my online existence, threatening my livelihood and my inability to discern danger from safety. These experiences became incredibly common, almost an unfortunate, defining factor of our early years navigating through social media. Although we would try to take action to defend ourselves, it was no use. Dangerous figures continued to remain in our vicinity, whether that was on the duplicate accounts or new ones they created quickly. I grew older and slowly started understanding what had happened to me and so many of my peers. As children, we had the right to a safe and fun social media experience. All we wanted to do was enjoy this new interactive life with one another.

When I browse social media currently, it is unfortunate to see the same, and arguably heightened, behavior showcased towards other minors on these same platforms. Time and time again, I have seen concerns for the safety of children present on these platforms, to no avail. Instances of child sexual abuse materials (CSAM) circulating on these very apps, the luring and grooming of children, and sextortion continue to rise exponentially. So, how do these companies claim they have taken further measures and action to prevent this?

Discord: Your head of Trust and Safety, John Redgrave, said in an October 2023 interview that technologies for CSAM detection on your platform "could be extended to video with enough effort." I demand that this effort become more than just words. As the spread of CSAM grows, it is imperative that these detecting technologies also apply to videos and not just stills.

META: An unredacted version of the New Mexico lawsuit provided new information to the public regarding META's knowledge of just how extensive exploitative material on the platform was. Another acknowledgment was of the popularity of your app with minors as young as six years old and how Meta "leveraged that" to achieve the goal of Facebook Messenger becoming the most popular for those audiences by 2022. But, META employees also flagged "sex talks" as 38 times more common on Facebook's Messenger. With this in mind, the platform must consider the popularity of its app with children and commit to more than just acknowledging the existence of these issues on its platform.

Snapchat: Snapchat is among one of the most named platforms for CSAM. In 2021, the Vice President of Snapchat Global Public Policy, Jennifer Stout, claimed that "Snapchat was the anecdote to social media" as it "focuses on connecting people who already know each other" and "focuses on privacy by making photos and messages delete by default." This is incredibly flawed logic when approaching the issue of rampant child sexual exploitation and predators' access to children on Snapchat. It is incredibly easy, then, for predators to establish and "sustain" relationships they have with children, and to send explicit content secretly with said default delete. So, how will Snapchat respond to this increased risk?

TikTok: TikTok claims to have a "zero tolerance policy" for CSAM on its platform. But, according to numerous investigations by Forbes, TikTok's "post-in-private-accounts" feature makes it easy for predators to meet and send sexually explicit images to children. Other features, like an easy workaround for banned accounts, showcase the weak spots in this zero-tolerance policy. As more minors continue to gain access to TikTok, how will the platform further advance its efforts in safeguarding children from predators? TikTok must implement better technology to monitor accounts, photos, and video content effectively.

X (Twitter): According to X's transparency reports in 2022, they have been unwilling to cooperate with their own transparency mandates. This runs concurrent to issues like a decreased trust and security team. In the fall of 2022, when Elon Musk acquired X, and after he stated "removing child exploitation was his number one priority," the team responsible for regulating and reporting CSAM on X was cut from 20 to 10 people. Furthermore, Musk has reportedly disbanded X's "Trust and Safety" Council, a group of volunteers who would offer the company advice about online safety. Issues surrounding X's database PhotoDNA, which reportedly

ENDOSEAC
COALITION TO PROTECT KIDS ONLINE

detected and flagged CSAM, had ineffectiveness issues, as accounts flagged for CSAM were still up and running. Stanford's Internet Observatory also reported 128 accounts advertising the sale of self-generated CSAM, and although most accounts were taken down, "a reported 22 of the 128 were still active over a month later." Will X commit to furthering its efforts to overcome these hurdles and make detecting, reporting, and taking down these dangerous accounts easier?

I ask that these platforms further evaluate the effectiveness of their current technology when detecting CSAM and other predatory behaviors on their platforms. The present safety features are just not sufficient. I also ask that they listen to survivors and a very concerned society. This request can no longer be merely "considered" and once again neglected. Social media is only continuing to advance and grow, and it is your absolute responsibility to ensure the safety and security of your users. I do not want another child's experience on social media to be one marked by a lack of safety and risk of exploitation. You have the capability to make these changes efficiently, and I demand that you do so to ensure a safer, more equitable future for social media users, especially those who are children and survivors.

Thank you,

Christine Victoria Almadjian



Statement for the Record by Vanessa Bautista, Global Survivor Network
and Philippines Survivor Network
Senate Judiciary Committee
"*Big Tech and the Online Child Sexual Exploitation in Crisis*"
January 31, 2024

Chairman Durbin, Ranking Member Graham, and distinguished members of the committee, thank you for the opportunity to submit a written statement on behalf of the Global Survivor Network on the issue of combatting Online Sexual Exploitation of Children globally.

My name is Vanessa Bautista. I am an American citizen with Filipino ethnicity. I am a survivor of child sexual abuse. I am a Founding Member of the Global Survivor Network (GSN), an international group of survivors shaping and leading a movement to protect people from violence. With over 4,000 members in ten countries, the GSN is leading a movement to protect communities by empowering and mobilizing survivors of various forms of violence, including human trafficking, police abuse of power, violence against women and children, and online sexual exploitation of children. The GSN started in 2019 and draws from the learnings and experiences of survivors globally. We strongly believe that survivors are experts on the issues of violence and discrimination that we have suffered and that our collective voices will inspire change[1].

One of these GSN groups is the Philippines Survivor Network, based in regions across the Philippines. "The PSN, Philippine Survivor Network, aims to contribute to creating safer communities in the Philippines, in which vulnerable people are protected from all types of violence, the survivors are sustainably restored, and the justice system listens and works together with survivors" said Monica Renomeron, PSN core-member.

Online Sexual Exploitation of Children and Child Sexual Abuse Materials

The internet has profoundly connected people worldwide, transcending borders, enabling instant communication. Through social media, email, and messaging apps, individuals maintain relationships irrespective of physical distances. The internet serves as a platform for sharing information, ideas, and knowledge, fostering a global community. There is, however, a dark side to this: a side where children reap the violent outcomes of the internet. Children around the world suffer abuse at the click of a mouse. Perpetrators log online, and abuse children over livestream, with thousands of onlookers. One might suggest the answer is cracking down on madams or pimps, rescuing children to set them free from abuse. However, even after the kids have been rescued, content of their abuse continues to live online, memorialized by the internet, for anyone to access.

As a Filipino American, and a survivor, I have personally experienced abuse as have the members of the Philippines Survivor Network group, some who were exploited online from incredibly young ages. One of my survivor colleagues was only 7 years old when she was trapped and forced

[1] Learn more about the Global Survivor Network at https://globalsurvivornetwork.org.

to sit in front of a camera while pedophiles from countries around the world: the US, Canada, Germany, etc., forced her to perform for them. I shared her story when I testified before Congress in 2023, at the House Judiciary Committee Subcommittee on Crime and Federal Government Surveillance "Children are Not for Sale: Examining the Threat of Exploitation of Children in the U.S. and Abroad."[2] Today, Joy[3] joins me in endorsing this statement along with other survivors from the Philippines Survivor Network. Many of the survivors in the PSN, have been victims of this horrific crime. In 2023, International Justice Mission and the University of Nottingham Rights Lab released an unprecedented report [4]on the trafficking of children to produce child sexual abuse material (CSAM) in the Philippines. It revealed that 471,416 Filipino children were subjected to this crime last year – 1 in 100 children.[5]

1 in 100 children[6] should have been protected from this crime; They should have been able to enjoy playing with their friends and going to school without pain and trauma from experiencing this dark crime. Survivors want justice, and safety, and are here to work alongside stakeholders to ensure our communities and children are protected around the world. Survivors also want to be fully free themselves. This means not being restricted by fear and living in constant anxiety because their content is on the internet. I urge Congress to prioritize removing Child Sexual Abuse Materials online. This is not merely pornographic content, these are violent, graphic, painful, livestreamed abuse of young boys and girls available for consumption by anyone, with just a click. Survivors deserve privacy and freedom, and we must create spaces that enable them to live their lives with dignity and true freedom.

To address this issue, it is crucial to recognize that solving this problem requires collaboration beyond the Philippines. We acknowledge that the government of the Philippines has made substantial efforts to investigate child trafficking for sexual exploitation, ensuring victims' safety, and holding traffickers accountable. However, the immense scale of online sexual exploitation of children (OSEC) material surpasses their endeavors. This crime thrives globally, with pedophiles contributing to its exponential growth. Combating such a global issue requires a united effort, with the US leading the movement to combat OSEC. Our recommendations are as follows:

Survivor Collaboration: I want to encourage us to keep survivor experiences at the forefront. It is important to highlight lived experiences in addition to seeking survivor input in designing policy. In 2023, the Philippines Survivor Network and the Inter-Agency Council Against Trafficking, a government body responsible for coordinating and monitoring all anti-human trafficking efforts of

---

[2] https://judiciary.house.gov/committee-activity/hearings/children-are-not-sale-examining-threat-exploitation-children-us-and

[3] pseudonym used

[4] Scale of Harm Report Summary,
https://assets.ijm.app/IJM_Scale_of_Harm_Summary_Report_Sept_2023_f733d4e011.pdf

[5] Scale of Harm Report Summary,
https://assets.ijm.app/IJM_Scale_of_Harm_Summary_Report_Sept_2023_f733d4e011.pdf

[6] Scale of Harm Report Summary,
https://assets.ijm.app/IJM_Scale_of_Harm_Summary_Report_Sept_2023_f733d4e011.pdf

the Philippine government, collaborated to draft a policy. The drafted policy was focused on adopting guidelines for engaging with survivors in policy development, program design, and implementation. This approach is critical because it goes beyond simply listening and understanding but also makes survivors part of co-designing solutions to the problems they have identified.

Financial Restitution: Survivors have been exploited for profit, and so it only makes sense that offenders must be held financially accountable. In April of this year, the Philippines Anti-money Laundering Commission released a study[7] of suspicious transaction reports involving payments for online sexual abuse of children. The top four countries in terms of offenders paying for online CSAM were the U.S. (by far), the U.K., Australia, and Canada. This is urgent: these offenders have taken years from survivors' lives. We must build a system where foreign survivors outside the US directly receive financial restitution to heal, recover and reclaim the lost years of their lives.

Removal of Child Sexual Abuse Materials: It is essential to remember that in this crime, an act of abuse is recorded and streamed digitally. Imagine the most vulnerable, painful, exploitative moments of your life are just one online search away. How can we expect survivors to carry on with their lives when their privacy is violated every day? In 2021, Apple promised child safety measures, and 2 years later, our survivor groups are still petitioning to remove images of abuse and protect children online. All CSAM should be prevented from being distributed online, and existing material must be removed. Please hold tech companies accountable when they fail to keep children safe on their platforms and incentivize them to do better in the future. The PSN had addressed the EU Parliament and European Council in 2023, urging them to take action to protect survivors.[8] We echo the key points of our statement to the US Congress, requesting
i. a balance of user privacy with child protection and the privacy of victims and survivors,
ii. Tech companies should be required to detect both new and known child sexual abuse material and finally,
iii. Tech companies should be required to prevent livestreamed child sexual abuse.

Endorsed by:

Philippine Survivor Network and Global Survivor Network

Philippine Survivor Network

GLOBAL SURVIVOR NETWORK

[7] http://www.amlc.gov.ph/16-news-and-announcements/456-updated-list-of-dnfbps-as-of-30-june-2023
[8] Read the full letter here
https://www.ijm.org.ph/assets/resource/PSN_Letter_to_EU_Parliament_and_Union_Council.pdf

Home Office

Chloe Squires
Director General
Homeland Security / SRO CT

2 Marsham Street,
London SW1P 4DF
DGHomelandSecurity@homeoffice.gov.uk
www.homeoffice.gov.uk

The Hon. Dick Durbin
Chairman
Senate Committee on the Judiciary
United States Senate

8th February 2023

Dear Chairman Durbin,

UK evidence in support of Senate Judiciary Committee with Big Tech CEOs on Child Safety

Thank you for requesting the UK's position on child online safety to inform the Committee's consideration on this important topic. On behalf of the Home Secretary, please find attached two documents setting out the UK's Online Safety Act 2002 and our position on end-to-end encryption (E2EE). I hope these assist in providing context of our own legislative framework and balanced approach to security and privacy concerns to ensure public safety.

We remain committed to working with you and the Committee on this important issue and striving, where possible, to be aligned in our approaches and principles. The Home Secretary is keen to discuss this further in his upcoming visit to the US, where we can discuss next steps and further areas for collaboration and support.

Yours sincerely,

Chloe Squires

Director General Homeland Security and SRO CT

February 5, 2024

Dear Chairman Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

I am a survivor of child sex trafficking and child sexual abuse material (CSAM). The sexual abuse perpetrated against me began when I was six years old. I was in first grade. I really loved grape soda and coloring. Photographs and videos of my abuse were initially taken to direct me as to how I could "do better" and "be a good girl." I was made to rewatch and analyze the imagery, receiving instruction around where I could be more compliant or less depending on what an adult asked for. I was trafficked for the first time when I was seven years old. I still loved coloring, but I had stopped speaking. The strangers I was sold to were called "clients." Clients would pay extra to have photographs and/or video*s* taken while they raped and tortured me. I watched clients hand over payment and in return, they owned me for a time and the CSAM forever. I was sold to hundreds of abusers with countless photographs and videos created over the next eleven years throughout the East Coast of United States and Canada. I grew up knowing that my worth was defined by the amount reflected in monetary transactions.

I watched the Senate Judiciary Committee hearing on Wednesday, January 31, and am writing to share my thoughts as a survivor of familial child sex trafficking and child sexual abuse material.

It is important to understand that the safety features and parental controls that tech CEOs tout as acceptable solutions to the epidemic of child exploitation on their platforms are rarely relevant for survivors whose abuse and imagery were orchestrated by their own family. My parents were my abusers, and they sold me to other abusers. This happens far more often than any of us really know because survivors of familial child sex trafficking and related crimes like CSAM are less able to report the crimes perpetrated against them and less likely to receive interventions while being victimized. I was never in ownership of the CSAM made during my abuse so tools like Take It Down do not serve me. Parental controls are not helpful if it is your parents who are abusing and exploiting you. How are we supposed to intervene for victims like me when doing the bare minimum is voluntary for companies who have proven themselves to prioritize profit over children?

I understand that The REPORT Act has passed through the Senate, and I am thankful for this important step toward requiring tech companies to report child sex trafficking to the National Center for Missing & Exploited Children. I am writing today to ask the Senate to advance the remaining four pieces of legislation that passed with bipartisan support through the Senate Judiciary Committee. Further, as mentioned numerous times in the Senate Judiciary Committee's hearing, I would like to see a Federal regulatory agency established to hold tech companies accountable to basic child protection measures like proactive detection and removal of known CSAM on fully encrypted platforms. If we allow tech companies to get away with offering safety features and parental controls as solutions to the egregious harms on their products, we are going to lose entire generations of child victims.

I would like to thank the Senate Judiciary Committee for their commitment to these issues. It is imperative that the Senate act with urgency for all survivor populations, including those victimized and exploited by their own families. Please do not forget us when considering what actions will be taken moving forward.

Thank you for your time.
Abigail, survivor of child sex trafficking and CSAM

February 5, 2024

Dear Chairman Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

My name is DJ. I am a 29 year old male survivor of child sex abuse and production of child pornography (CSAM). I still discover things on a daily basis that the abuse and exploitation I have suffered from has hurt me. It set my life on the wrong track, and destroyed the normal childhood, teenage years, and early adulthood, including a marriage that involves 3 young children. The abuse all started when I was only 10 years old. I constantly heard the common sayings: I was special, he cared about me, and that we had our own "special secrets." Most of the time I spent with Joel, I did not have clothes on or very little if I did. There were constantly pictures and videos being taken. After the abuse, on the verge of tears falling, he would try to get my mind off things by playing video games or turning on a movie.

To this day, the games we played, and movies we watch are constant reminders of my very traumatic past. It seems like every day there is something that reminds me of this past. This continues to be extremely hard on my mental wellbeing. I received multiple messages from Joel wanting to reconnect and "rekindle" our relationship. Suicidal thoughts sky rocketed from the stress and hurt from the past as it felt there was not an escape.

Every day I live in constant fear that someone will see my pictures and/or videos and recognize me and that I will be humiliated all over again. It hurts more than words can describe that someone has looked at pictures of me as a little kid being abused to get some kind of sick enjoyment from it.

What happened to me hasn't gone away and it never will. This affects me and survivors every single day.

Joel hid behind the screen names:
Dragual - World of Warcraft
Official Foxy - Facebook
Officialfox2017 - Instagram and X
Official.Fox - Snapchat

He continued to use these platforms as well as countless others to groom and entice minors while he recorded these interactions.

As of this writing, 4 of the 5 listed accounts above, are still active today even after he has been convicted and sentenced to 40 years prison.

I could not have done this without my team. This deeply affected not only myself but my mom, dad, sister, grandparents, as well as other family and friends. I had a huge support system throughout this whole process that made this as easy as possible. I can honestly call my team the "Dream Team" which would include all the officers, investigators, attorneys, judges, and counselors.

I was blessed with the opportunity to be present at the Senate Judiciary Committee hearing on Wednesday, January 31st, and to be able to speak at the press conference following. I am writing to share my thoughts as a survivor.

- I feel that there is a huge swing of momentum going the correct direction as we were walking out for the day. However, I do not feel that Tech is any closer on approving or assisting us in passing this legislation. As Senator Graham stated, we will die waiting if we wait for them. I personally thought that the responses from the tech CEO were subpar at best. It was clear that they were prepared not to answer any questions.
- I feel the interactions with Senator Cruz, Hawley and Blackburn were most impactful over all. Some may see it as being overly aggressive. I feel wholeheartedly different. They need to know that we are not waiting any longer. We have had the conversations, we have kicked the can down the road waiting for them to do something. The time is now! I feel that was very clear and they were put on notice.
- As a survivor I want to thank each and every single one of you and your staffers for all the love and support. You are ALL heroes in my book. I am beyond honored for the opportunity to be present at the hearing. There was clear bipartisan support and how important this issue really is. As a survivor, it is so important to not feel alone. I truly felt that. This isn't just my fight, it's OUR fight. I would love the opportunity to be able to speak to each of you and to be able to thank you personally. It is so hard to put into words what this day meant. I pledge my support to each of you, I would love to assist with needs that may come up.
- Do we have a date or timeline for when this legislation may be voted on?

I am writing to ask that the Senate keep up the pressure to pass all of the much needed child protection legislation. If it was not for all of your work, we wouldn't be where we are today and wouldn't have been able to put them on blast as you did. We have to keep holding tech companies accountable for their actions and continue to demand answers and results. Something they have continued to refuse to do. You see us, you hear us, you are there for us.

I again want to thank you for the opportunity to be at the hearing. This was the most rewarding experience that I have gotten to be part of as a survivor. I have so much hope and drive to continue this important work. I want to take this time to pass my thanks from all of my family, friends, co-workers and community. I am currently working on getting a meeting set up with Senator Grassley to discuss this topic on both a local and national level.

This day will go down in history! We, together, will continue to make a difference. I cannot wait to watch this continued progress and for these bills to become law.

Thank you!
Dj
Survivor

Hi, my name is Elle, and I'm now 20 years old. I am a student at a large university studying Psychological Sciences and Biochemistry with the hopes of becoming a Physician Assistant when I graduate college.

When I was 15, I was exploited by a serial online predator from the United Kingdom. He manipulated me into sending thousands of images and videos in order to profit off of my exploitation. After about a week of constant harassment, I could not take it anymore and cut all contact to escape the torture. Because I ignored and blocked him, he released images and videos of me to the dark web, which then ended up on platforms like Facebook, YouTube, Twitter, Reddit, Instagram, and pornography websites. These child abuse materials were duplicated and distributed thousands of times before law enforcement could take some of it down. There would be days where I would report the content and it would take weeks before anything was done, if at all. Content is still being shared to this day.

After the content was posted to social media, I deleted all of my accounts, left my school, and moved to an online schooling platform. I could not bear to face everyone at my school, many of whom saw the images of my victimization. I lost a lot of friends during this time, and I hardly left my parents' house for about a year. I never felt safe in my house because I was reminded of what happened, but I also could not leave in fear that I would be recognized. I was also receiving messages from strangers saying that I was "disgusting," a "whore," that I should be "ashamed of myself," and that I "should kill myself." I took this to heart.

Once I made the decision to leave school, my mom quit her job to stay home with me and make sure that I didn't harm myself. She would sleep with me every night and we would read together. My relationship with my siblings was significantly harmed; my sister left home prematurely and my brother and I stopped talking. A month after I left school, I decided to go to therapy. I dyed my hair and got different colored glasses to avoid being recognized. I remember feeling isolated because my therapist couldn't help me enough, or really at all. I couldn't look at certain innocuous objects because they would remind me of images the predator sent me of other children being abused to convince me that what was happening was normal.

My inability to be on social media makes me feel disconnected from my peers, and people in school have questioned why I am not on social media. At school, I struggle to meet new people, anxious that they know about my victimization and are judging me. I also keep myself busy to avoid thinking about the past, by spending much of my time doing schoolwork, involving myself in clubs, internships, and going to the gym. With the friends that I do have, I'm constantly making sure that I do not upset anyone because I am afraid that my friends will leave me if I do something wrong.

Several years after my initial abuse, I experienced an incident where a stranger reached out to my place of work at school asking if I worked there. My workplace called me asking if I knew this man, and I eventually had no choice but to explain to them what happened to me when I was 15. This stranger then tried to contact me through my school email saying, "I have seen stuff circulating of you and would like to see if there's an opportunity." This was followed by "I've got

some vids of you doing some strange stuff, wanna talk about it?" And lastly, "Imagine if these vids got into the wrong hands, would ruin a career and have prison time." I don't know how he got my information. I have to live with the fact that my name and child abuse materials are out there, traded among predators.

The National Center for Missing and Exploited Children emailed my mother in March of 2023: "We do routinely search her name to try to find this content and notify the company of its existence, so know that we are here right alongside with you trying to get the content removed as quickly as we can." I didn't realize that they were still removing content, as I avoided looking myself up to protect myself from spiraling. It upsets me and makes me feel like I'm reliving the trauma when I see images of my victimization online, especially when I know there is a team constantly taking content down.

As for my future, I don't like to plan too far ahead because I know how quickly your life can change. This event and the constant reminders of it have influenced my decision to not go to medical school, to not have children, and to not live anywhere near where I grew up. I don't want to have to have my life taken away from me again or harm others that are associated with me.

I just wish I could be a normal person again, and I don't know if I'll ever get that life back.

I am writing to ask the Senate to not let what happened to me happen to any more children. Do not wait to pass the much needed child protection legislation. Do not hesitate, because lives are at stake. Big tech does not currently have enough protections in place to protect children who are victimized in this way. It should not be an option for application users to see content that has been flagged as potentially harmful. As someone who has reported CSAM images to these companies, and have been told that they do not violate their terms of service, this makes me feel as though I don't own the rights to my body and content. In addition, if my name is being used as a hashtag to easily find my abuse material, big tech companies should flag these hashtags and related search terms to prevent them from being posted and shared. Big tech companies should be doing more to protect victims as well as prevent the sharing of material to predators.

Given the current advancements in AI technology and facial recognition, there should be an option for survivors to opt out of having their images on the internet. If a survivor is reporting one picture, a company should be able to connect their face to other images on the internet and proactively remove those. It is important for the rights of the victim to be prioritized over the rights of the content sharer, and if I ask to remove images of myself, it should be automatically taken down. If I had the option to opt out of my images being on the internet, or more proactive removal of images, I can't imagine how different my life would be. I want to be able to control my image and not be revictimized by big tech companies failing to do their job.

If companies had these additional protections in place, then maybe I would be able to be active on social media like my peers. Unfortunately, until that happens, I can't be a normal kid because of the risks that I know are out there.

Thank you for giving me the opportunity to share my story, and I pray that you don't have to listen to any others.

Sincerely,
Elle
A Survivor

Leah Juliett
leah@leahjuliett.com

January 31, 2024

Distinguished Members of the Senate Judiciary Committee,

When I was fifteen years old, four images of my naked body were uploaded to Facebook.

After being disseminated like trading cards on Facebook Messenger, they were uploaded to a worldwide image board, which is still accessible behind a paywall today.

The man who initially exploited my body is currently incarcerated.

But the man who allowed my abuse to spread on Facebook was in that hearing moments ago. His name is Mark Zuckerberg.

As the CEO of Meta, Mark Zuckerberg has the responsibility to prevent the abuse that happened to me on his platform from happening to any other child or teenager in America. Because it didn't, and it isn't just happening to me.

The negligence of Mark Zuckerberg and our the other tech CEOs that we heard from today has led to a widespread epidemic of hate, exploitation, and child sexual abuse on the Internet. I know because it happened to me. The lifelong trauma that comes with exploitation like this cannot be overstated. It tried to steal my body, my dignity, and my freedom.

But I must be clear: the internet that tried to kill me is the same internet that saved my life.

Safe online spaces are lifesaving and life-affirming, especially for young queer and trans people navigating identity and self-expression. Instead of restricting young people from accessing these platforms—we must make these platforms better and safer for the young people who use them.

To our legislators and tech CEOs—survivors everywhere are watching your next move. This legislative cycle, I want to see technology companies continue to be held accountable for their negligence. I want to see them enact policies that directly address the harm their platforms are causing.

Survivors are our greatest teachers. We are technology's most profound resource. We continue to stick out our necks and tell our stories in the face of peril, in the face of our most ardent abusers, demanding justice, even those of us who will never see our day in court.

Today is that day. The court is the United States Capitol, and our justice is this moment.

My name is Leah Juliett. I am a survivor of child sexual exploitation, and today, I am demanding a better online world than the one that I grew up in.

February _6_, 2024

Dear Chairman Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

I am a survivor of the production of Child Sexual Abuse Material in the early 2000s in the Midwest from ages 10-12. You met me in the opening video, my name is Lexie. Section 230 had just gone into law a few years prior to the torture I endured, and it did nothing to protect me or the other children being exploited by a teen in the cul de sac of my grandmother's neighborhood. Since elementary school, I have been living in a social prison isolated by the reality of that content living on the internet while my perpetrators live freely to commit the crime again. My experience made me the advocate that I am today for other survivors, working modern cases where people never have to be in physical proximity of a child to isolate and exploit them at an exponential rate.

I watched the Senate Judiciary Committee hearing on Wednesday, January 31, and am writing to share my thoughts as a *survivor, advocate, and cybersecurity professional:*

- *Watching Senator Hawley demand that Mark Zuckerburg issue a public apology says it all. Big tech must be regulated into ethical compliance if we ever wish to actually address the privilege sextortionists are exploiting within their platforms.*
- *With the context of most of those companies rushing to implement changes one week before the hearing shows that they do know what to do, how to do it, and have the money, talent, and capacity to do so but they simply won't unless there is immense pressure to.*
- *Thank you for holding the hearing. I hope that it will not be the last one in order to maintain momentum, and I hope that in the future perhaps you could call on experts on algorithms, experts that assess the impact of these crimes both on an individual but also our systems as a whole, law enforcement who work these cases, and survivors who are also experts and have ideas on how to address the problem.*
- *Restitution was brought up, and I think that is a really important part of motivating tech and media companies to make the changes necessary to prevent and respond to incidents on their platforms.*

I am writing to ask that the Senate see to it that not only is Section 230 amended with already proposed legislature, but that it is extremely clear, so we don't run into the issue of a difference in interpretation, which is how tech and media companies get away with aiding perpetrators. The ability for victims to seek restitution is also important as a big incentive for them to do the right thing, and necessary for the recovery of victims. Personally, it has cost me well over $100k to recover from what happened to me, and I will continue paying for damage from the transgressions of my exploiters and rapists. Only 2% of cases ever make it to trial, which is the requirement to access victim funds. It's my understanding that there's consideration to slash those funds, which would be a telltale sign to survivors and perpetrators always watching, that these acts have become an acceptable part of life. In the future, I hope there will be bills to protect and make victim funds more accessible. Additionally, I hope that the Senate will

look into why the PROTECT Our Children Act of 2008 funds were never deployed to the incredibly brave men and women who investigate these crimes. They desperately need funding to access the technology they need to more efficiently work through their enormous caseloads.

Thank you for allowing me to participate by submitting a statement. I hope that I can count on you to see this issue all the way through and in every area that I've mentioned. For the future of every child in America, I beg you to keep your foot on the gas. You have an army of survivors, advocates, nonprofit organizations, law enforcement, faith groups, and families grateful to be heard and hopeful to not be disappointed.

Thank you.

Lexie Smith, Survivor

February 2024

Dear Chairman Durbin, Ranking Member Graham, and Members of the Senate Judiciary committee,

My name is Millie. I am a survivor of sextortion and CSAM. I began being sextorted when I was only 16 years old. My victimization began when I received a message on Facebook from an unknow account saying that they had images of me and if I didn't comply with what they wanted from me, they would send the images out to my friends and family. I thought this would be a one-and-done deal. Little did I know that I would continue to be sextorted for the next 16 months of my life.

After my mother discovered what was happing to me she reported it to the police and my abuser began distributing all the images and videos I had sent him. He created multiple Facebook accounts befriending all my friends, family, and community and posted hundreds of nude or partially nude images of me. Facebook did nothing to support me or help me during this time. In fact, it wasn't until the FBI took over the case and threatened legal action against them, that Facebook finally complied in shutting down the accounts and removing the images. But by this time so much damage had already been done. My images had been shared and viewed over a million times. My childhood innocence had been ripped out of hands and I would NEVER get it back.

After watching the hearing, I want to thank you for asking some insightful questions and proving lots of information about what these companies are doing and not doing to protect children. I think we need to keep the pressure to pass the package of child protection bills. If we don't protect our children then who will?

Many people like to say it's the responsibility of the parents to keep this from happening to our children. However, I know from personal experience that my mom did everything she could to protect me. She monitored all my social media accounts, checked my phone regularly but she couldn't do this 24/7. It only took one little moment for my victimization to begin. We have to start making these big tech companies accountable for what their platforms are doing to children. Providing a safe place for children on the internet shouldn't be this difficult. These companies claim they have age limits and are monitoring what children are seeing but they aren't doing near enough.

How many more children have to be sextorted, have distribution of CSAM, or take their own lives before we take some serious action?

I thank you for allowing me to submit a statement on this issue. Giving survivors an opportunity to chime in on issues that have affected and changed their lives is incredibly powerful.

Thank you,
Millie - A Survivor

February 06, 2024

Dear Chairman Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

I am a parent of a survivor, an innocent child.

I watched the Senate Judiciary Committee hearing on Wednesday, January 31, and am writing to share my thoughts as a parent.

- *My impressions of the hearing were thankful but to late for my child.*
- *Big tech companies are contributors to the problem.*
- *I feel that the smaller companies are willing to help but not so much for the bigger companies… Facebook in general.*
- *I want to thank you for holding the hearing. I feel you didn't go far enough in pressing the tech CEOs for answers. I appreciate you standing up for survivors/parents. When are you going to be able to reach consensus to vote on legislation, so we have new laws?*

I am writing to ask that the Senate… *Please listen to survivors and parents. Please keep up the pressure to pass much needed child protection legislation. Please keep holding tech companies to account. What is your next step?*

I want to thank the Committee for allowing me to participate in this forum. Please don't give in or up to Big Tech! I want to let you know the immense importance of not allowing this issue to fall by the wayside and to keep up the pressure, so we have some new laws this year.

Thank you.

James Jessee

Parent of a surviving child

February 6, 2024

Dear Chairman Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

My name is Julia McDonald. I am the mother of a CSAM survivor, as well as the Director of Team HOPE, which is a program of the National Center for Missing and Exploited Children. Team HOPE is a volunteer program made up of parents, family members, and survivors who reach out over the phone to offer emotional support to other parents, family members and survivors who have experienced missing and exploited children situations. In my work, I have spoken with thousands of people who have experienced having a child in their life be sexually exploited or are survivors of exploitation themselves.

I watched the Senate Judiciary Committee hearing on Wednesday, January 31, and am writing to share my thoughts as a mother and peer of other parents who have shared their experiences with me. I appreciated the way you would not take responses to yes or no questions that were filled with legal speak that tried to make it look like the CEO's would support the laws, when in fact they just wanted to have more time to try to lobby for laws that are more in their favor. The surprise for me was the CEO of Snap, who did seem to agree to the laws you have proposed.

What I have heard from countless parents and family members is that they thought they were doing everything they could to keep their children safe. When they find out that their child was victimized by a predator right in their own home, that is something that shakes their sense of security in their home to the core. We have created super computers that fit in the palm of your hand and have unleashed the potential for predators to reach children of all ages via social media. To do nothing about this is to say that this is okay to sacrifice the innocence of our children so that a handful of billionaires can make more money. I, for one, say not one more child.

I am writing to ask that the Senate pass the bills to protect children immediately. Please continue to listen to the families and survivors who have been adversely affected by the lawlessness of the social media frontiers. Children have died. Children have had lifetime adverse effects which require support and an understanding of the issues that being a survivor of sexually exploitation involves. How many more children are going to experience being victimized while the lawyers and lobbyists put a pause on meaningful laws for the sake of money? Please pass these bills now.

I appreciate your bi-partisan support of this issue and I hope you will get these bills passed into law by the end of the year. The children and their families cannot wait any longer. The time for action is now. Thank you for your hard work to bring more awareness of the issue of children's safety online. Please protect the children, not the CEO's.

Thank you so much,

Julia McDonald

February 6, 2024

Dear Chairman Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

My name is Zack, and I am writing this letter to you as a fellow American citizen who is, as I believe you are, concerned for the well-being of our nation's youth. I deliver this letter to you on behalf of my family, in hopes to share a piece of our story, and I write it particularly on behalf on my youngest sister who became a victim of online trafficking in 2010, at the age of only fifteen.

Fifteen. Please don't let that number pass you without impact, and please don't think it a typo or trivial. For those of us that need a reminder, fifteen is just barely a high-school sophomore. Fifteen is not old enough yet for a driver's permit, and it is far shy of the right to vote for anything more important than homecoming king or queen. Fifteen is in every way a child and yet, only a few dozen miles from our home, it was deemed plenty old enough for a predatory and disgustingly efficient system to take advantage of her. It was old enough for men, American men, to seek forced sexual encounters with her for pay, the profits of which were shared by the hosting website Backpage.com.

It took no special knowledge or skills or "dark web" to see the thinly-veiled advertisements for child sex trafficking hosted by Backpage.com, only a few simple clicks and intent. The website didn't even make any significant attempts to disguise their crimes, because in the eyes of American law and the protections offered by Section 230, they were completely immune to liability for their large part in the process. So it was that for a period of no less than 108 days, men ordered my sister with the same ease that someone might order a pizza on a Saturday night, and if that makes you feel sick to your stomach to hear as it does me all I can offer is this: we will not apologize for the truth, but we will ask for your continued help in preventing it from happening again to another child.

My family has fought in the over dozen years since to protect others from the harm that was dealt to my sister in various courts across the country, with support from many of your peers. Through tremendous bi-partisan effort, in 2018 the President signed the FOSTA bill into law, setting a critical precedent; that in America we decided that there is such a thing as a price too high to pay. That our children are not for sale. As much as that day meant to us, we know this fight is not done.

Having watched the recent Senate Judiciary Committee hearing on Wednesday, January 31, I was reminded just how true that statement is: this fight is not done. The boldness of internet media and tech companies has been earned overs decades of winning one-sided battles, hiding behind powerful legal protections and lobbyist groups, and controlling narratives through their own media pipelines. They have done this, all the while done this while becoming some of the most profitable businesses and individuals in the world, and yet still somehow, they can't find the money or sufficient technology to prevent the distribution of CSAM or the occurrence of trafficking in their own domains. It is straightforward greed, and it has won for far too long.

That said, it was incredibly empowering to witness our elected officials and government representatives take a stand for the victims of these terrible crimes, as you have done many times before. To stand apart from billionaire CEOs, apart from oppressive greed that would steal our innocent, and to demand that it be held these corporations take accountability for the active harm that they are causing to our future generations.

I am writing to you today to simply ask this; please do not stop. Continue to support important legislation such as the STOP CSAM Act, the REPORT Act, and the EARN IT Act, and others like them which have been vetted by the families of so many victims. We who are in this fight alongside you understand that real change is not fast. We understand that often it will take many, many failures to lead to one important success. We understand maybe best of all that very often the enemy is apathy, but that this fight is too important to put on the backburner while we take on easier challenges. You are fighting for the future of America, and you are not alone; and neither are we.

On behalf of my family to yours, thank you. Thank you for everything you are doing.

Signed, a loving brother.

Words on paper might never make a great impact, but I hope mine shift the conversation towards survivor-centered legislation and services. I was a victim in one of the largest sextortion cases in the United States which occurred when I was twelve. During this time, I already felt like I had no one to go to; no parents, no social support systems Naturally, I went to the internet as a refuge as most children/teens do today. I was targeted and had no idea how to even begin to ask for help. I was not the only one. Even outside of my case, **NCMEC received over 260,000 reports of sextortion in 2016. By 2021, those numbers had more than doubled**.

As you can imagine, if these numbers were on an upward trend in 2021 then logic dictates that they are still on the rise. This is evident by how many teen suicides we are seeing each day on the news due to financial sextortion. Children, the ones who we promise the future to, are ending their own lives due to bad actors who are seemingly protected by Section 230, lack of funding towards critical investigations, and not enough education on just how devastating sexploitation is. Adolescent mental health issues are already on the rise, as stated by the American Psychological Association in 2022.

The CDC-Kaiser Permanente adverse childhood experiences (ACE) study is one of the largest investigations of childhood abuse and neglect and household challenges and later-life health and well-being. The ACE's study investigates how traumatic experiences suffered in youth affect us as adults. For example, experiencing childhood sexual abuse raises the chances of adulthood cancer, obesity, and diabetes to name a few. Girls who experience childhood sexual abuse have a 2-13 times increased risk of sexual violence victimization in adulthood. I am not an exception to the rule. I suffered many suicide attempts throughout my childhood years. Now, I suffer with PTSD and other chronic physical health issues due to the online abuse and pure anxiety about my own safety and privacy.

Do we want to put a large majority of our future through a traumatic revolving door? I would sure hope not. We must speak out and enact real change. Our lives depend on it.

Like everyone else in this courtroom today, I deserved a normal childhood. We all deserve a childhood that we can look back on with fond memories and nostalgia. However, my adolescent years were instead filled with struggle and damage.

As a child, I was sexually, mentally, and verbally abused by my family and acquaintances from a very young age. I had a very rough life growing up. Yes, I did well in school, and I had some friends even though I found it hard to build relationships, and I had a somewhat positive outlook on life. [redacted] I was heading into 7th grade. A young child with dreams and aspirations that one day, I could find a way out of my terrible situation. I turned to the internet as a refuge, where I was finding new friends and reading about interesting topics for hours on end. That summer is when my initials would be written in a sexploitation indictment. I met a predator on the internet who stole what innocence I had left. I was given a threatening command to give a show on Skype, on the now defunct [redacted]

I was terrified – thinking that my life was in danger and that if I didn't comply something even worse would happen to me. My brain was nowhere near being fully developed – I hadn't even hit my teenage years yet. My body and mind simply went into survival mode, as it had many times before. Hour after grueling hour, I was given instructions as if I were a circus animal performing tricks on how to be "sexy" – taking my clothes off, parading my naked body around, inserting my fingers into my vagina to a point where I was in physical pain for months after the fact, and other things a child should never be doing for someone else over the internet. As the seconds passed and I was continuously pushing myself over the limit, I kept asking myself "when would this torture end?".

But that's the sad thing about trauma, it never simply 'goes away' or 'ends' as one would hope. After this terrible event occurred – my life became a living hell on earth. The battered and beaten girl I just told you about turned into an even more broken child. I began having more severe anxiety attacks, suicide attempts, and PTSD. I was constantly scared for my life. I couldn't trust anyone, and I automatically assumed that if anyone were to talk to me, they only wanted to get something out of me. I self isolated. I couldn't sleep, eat, or focus. With those feelings came a great bout of guilt and shame. I blamed myself, naive to the fact that what the defendant did was not my fault at all.

I felt so guilty that I didn't tell anyone – not even my family. I was left to hold this secret and just thought that if no one knew, it was better off that way. When I slowly started to dwindle into my self-harming behaviors and my long list of

suicide attempts, no one truly knew why I was doing it. In fact, I remember a loved one telling me that "it was just a phase" and that I would ultimately grow out of it. To this day, I still have not grown out of it, unfortunately.

[redacted] the FBI came to our house for the initial interview, I was disowned and looked down upon. I was told [redacted] "you shouldn't have been doing the things you were doing." And I was even grounded for months on end after – not even being allowed to go outside and be a kid the way I should have been able to. Eventually, word spread throughout the family and from that day forward I was looked at as a sexual deviant – when in fact – I was the victim. The defendant is the one to blame, while I have had to carry this flag and support myself on my own.

What glimmer of hope I had to have a normal childhood was stolen from me because of this. I no longer had friends or a positive outlook on life. As a direct result of this heinous crime, I spent seven months in inpatient mental health facilities due to numerous suicide attempts over the span of 8 years - one of which, I was almost successful. [redacted]

I cannot hold down a steady job because my anxiety is so bad, and most nights I cannot sleep well because of PTSD, nightmares, and tremors, and sometimes even paranoid delusions and hallucinations that I get when I am overwhelmed. Most of my hallucinations stem from not feeling safe, and that someone is out there to hurt me; or that they have the photos that I had to identify myself in when I was 13 years old. I fear that I cannot function as a good citizen of this society because I feel like I have to constantly look behind my back or delete my social media because someone else will try to do this to me again, or that when I speak, no one will take me seriously. Because he did not take me seriously. He saw me as a target and ran with it – intimidating me and pushing me into a corner that I could not pull myself out of.

Out of all my traumatic experiences, this one sticks with me the most. Maybe because for so long I could never put a face to a name, or even know who the defendant was. I felt as if I had raped my own body at the beckoning call of a pedophile. To this day I still feel so dirty, as if I had taken a million showers and

the dirt wouldn't come off my body. The defendant, quite frankly, has ruined me. Sitting behind a computer screen to anonymously blackmail children into doing explicit things is inexcusable. I cannot wrap my head around why someone would have such a motive to commit such a heinous crime.

[redacted]

happen to me, and not just to me, but to others as well. It tears me apart because I know that no matter what I do, these memories will stick with me forever. There are so many of us who have fallen victim to the defendant – most of which are not in this courtroom today. I simply cannot forgive, nor will I forget.

[redacted]

L.



February 6, 2024

The Honorable Dick Durbin, Chair
The Honorable Lindsay Graham, Ranking Member
Honorable Members
The United State Committee on the Judiciary

Dear Chair Durbin, Ranking Member Graham, and Honorable Senators:

ParentsTogether is a parent-led organization with more than 3 million members across the United States. We work directly with parents to address the most pressing issues and needs facing American families today, a primary one of which is online safety. Social media has become an urgent, daily crisis for millions of American parents and kids, and families are desperate for Congress to pass legislation to regulate social media companies.

Recent polling data from Hart Research Associates shows that a large majority of both voters and parents want to see Congress regulate social media, and that support is consistent across political parties. Both voters and parents think social media does more harm than good for society overall, and especially for kids and teens.

Behind those trends are real families – real parents and real children, being hurt really badly. ParentsTogether collected experiences from 23 parents from across the U.S. whose children suffered death or serious harm from social media. Their experiences represent the wide diversity of families experiencing the devastation of an unchecked Big Tech industry and the breadth and depth of the ways social media can destroy kids' health and safety.

We strongly urge Congress to act quickly and pass online safety legislation to protect children and prevent more families from experiencing what these parents have experienced.

ParentsTogether



Parent Name: Joann Bogard
Child Name: Mason Bogard
Evansville, IN
My child was harmed on YouTube in 2020

I was the parent who had all of the protections on devices and had the hard conversations, but that wasn't enough to protect Mason from algorithms feeding him the viral social media challenge called the choking challenge / blackout challenge. Mason had the false impression that this challenge was safe, but it went horribly wrong and he died.

Every week I retraumatize myself by searching, finding and reporting many choking challenge videos that I find across all of the platforms. Their Terms of Agreement states

that they do not allow and will remove all challenges from their platforms. I report them using the system that they have in place. They are very seldom removed. This is blatant and they know that they can hide behind Section 230 and claim immunity. Parents can't fight alone and Congress must step in and disallow this business model.

Parent Name: Rosemarie Maneri
Child Name: Shylynn M. Dixon
Lisbon, NY
My child was harmed on Facebook in 2019

My daughter took her own life due to sextortion on Facebook. They need to be accountable for their actions. Someone needs to keep our kids safe from online predators. Nothing will bring my child back but social media needs to be held accountable so we do not continue to lose our kids.



Parent Name: Neveen Radwan
Child Name: Mariam
San Jose, CA
My child was harmed on TikTok and Instagram in 2010

My daughter was targeted by the algorithms on Instagram and TikTok and sucked into the eating disorder “black hole” online. Rather than seeing the exercise and “healthy eating” content she was searching for, she was shown dangerous content consisting of “what i eat in a day” videos of people eating less than 500 cals/day or dangerous challenges of how to prove you are skinny.

No child or family should have to survive the trauma our family did because a company made MONEY by showing our child content they never should have seen and nearly died for it. No family should have to lose years of their child’s life or their family’s life together so a company can turn a profit.



Parent Name: Judy Rogg
Child Name: Erik Robinson
Santa Monica, CA
My child was harmed on Facebook and YouTube in 2022

My son, Erik Robinson, died from a viral internet challenge, The Choking Game, in April, 2010. Back then we didn't really "connect the dots" although in 2012 ABC's Nightline did an expose featuring Erik's death and the proliferation of "how to" videos on YouTube. YouTube declined to comment. It is only within the past few years since whistleblower testimony came forth that we understood the connection.

Social Media is the only business that is entirely unregulated for safety concerns regarding children. No other industry would ever get away with what Social Media companies get away with. It is simply un-American to refuse to keep our kiddos safe.



Parent Name: Tracy Kemp
Child Name: Brayden
Portales, NM
My child was harmed on Instagram and SnapChat in 2022

My son was cyber bullied on multiple platforms for being Black. During his 8th grade year in Lubbock, TX and Instagram page was created called ""Monkeys of Laura Bush Middle School"" with the tagline reading ""Send pictures of monkeys"" there were only

Black students on this page. My son did not even have an IG at the time, another student sent him a screenshot that he later showed us. After a day of constantly reporting the page it was finally removed. I reached out to IG to see who was behind the page and to this day never got a response.

Later a SnapChat was created under a similar name that begin to harass my son. After going back and forth for a few days he finally let his dad and I know what was going on. We tried to report the page and find out who was behind it but like most things on snap chat it ""vanished""

The emotional trauma and after effects of these incidents strongly impacted our family. My son was depressed, he wasn't himself but most importantly he couldn't understand why he was being bullied for simply being himself. It made our two younger children uneasy. As Parents we felt helpless and alone. It was a very difficult time for our family, even extended family members felt the effects of this happening to their grandson, and nephew. Protect children! We need more studies and education on how this affects our youth. We need to listen to these stories and understand the MANY harms that are out there waiting for our children. When it comes to social media we need to look at it like a product that is being sold. That product has been deemed harmful to children, and like other things that are harmful we do a recall. We let the public know what the harm is, not to use it until ____ and we also hold the companies responsible for not only fixing the problem but cannot put that product out until then. That is what needs to happen for social media.

We can't ""recall"" social media, but we can study the effects on children, we can hold big tech accountable with the duty of care, we can ensure the parental settings are easy to find, use and adjust. We need more moderation, ensure that who's checking to see if the content is appropriate knows what to look for, when post/pages/content is flagged that a human is checking into it and its not just taken down for a day or so. Mandatory investigations into minor deaths that occurred because of social media . Most importantly REGULATIONS!"

Parent Name: Andrea Rossi
Child Name: Julia
Farmington Hills, MI
My child was harmed on TikTok, Instagram and SnapChat in 2021

My child was not allowed to have social media until graduation from 8th grade. Within 4 months of being on online platforms, she developed anorexia nervosa. She noticed pictures of several girls which were clearly filtered and felt she was too chubby to compete with other peers. She started to click on multiple videos that had to do with intermittent fasting, calories apps, "what I eat in a day", and "cleanses". As she clicked on these, more and more of her feed started to be nearly 100% diet culture related. She was 14 at the time.

She still struggles to this day with anorexia and still, even with filters in place, videos pertaining to weight loss, exercise, and diet fads leak into her feeds. Diets and fasting are NOT meant for children. I found 8 different tracking apps she downloaded from these SM platforms pertaining to fasting, exercise, and calorie counting when she was acutely hospitalized for malnutrition/semi-starvation with a heart rate in the low 20s. She could have dropped dead at any minute. While anorexia is a complex disease, the feeds she was getting on SM contributed to how quickly she made a downfall. There should be stipulations in place for kids under 18 to have access to diet fads and misinformation that are not from health professionals on diet/exercise/skincare/medical advice. Literally ANYONE can make claims on these platforms and kids believe it is legit information from a good source.

Parent Name: Carmen Mouritsen
Child Name: Nigel
Cottonwood, AZ
My child was harmed on TikTok and SnapChat in 2017

On April 8, 2021 the we lost our beloved son to the viral choke out challenge. The destruction caused by his senseless death is indescribable. The pain is unbearable. We just want him back, but there's no going back. It takes one bad decision and it's over. He was only 14. He was so loved, so loving and funny and smart.

April 8th was a wonderful, normal day. He went to school, we ate dinner, we laughed, everyone was happy. Then I heard the screams. His sister found him hanging from the shower curtain rod. There were marks on his neck where he tried to free himself.

Our hearts and family will always be broken, there is no such thing as getting over losing your baby. Please help. I don't want anyone else to have to live with this kind of horror. I would like dangerous content to be removed immediately. I would like parents

to be alerted when children are visiting dangerous sites or viewing inappropriate content. I would like it to be illegal for algorithms to target children.



Parent Name: Sharon Winkler
Child Name: Alex
Atlants, GA
My child was harmed in 2022

My son, Alex, was 17 years old when he died by suicide after he was influenced by anonymous users online. Alex was a fun, sensitive kid who was active in Boy Scouts, theater, band and his church youth group. He had good grades in school. It was, and still is, inconceivable to me that pro-suicide content, clearly against platform policies, remains online.

Deaths by suicide can cause shame and guilt for parents. As I have become more open about how Alex died, I am asked to contact parents whose children have also died by suicide. Many tell me stories of how their children were sextorted, cyberbullied and algorithmically bombarded with content that exacerbated their child's depression and anxiety. Technology companies need to design products with safety in mind, so that these tragedies can be averted.Technology companies have clearly failed to regulate themselves. Federal regulation is necessary to ensure safer experiences for our children and teens. The Kids Online Safety Act (KOSA) of 2023 would require covered online platforms to provide safer online experiences by:

• Imposing a "duty of care" on online platforms to prevent and mitigate specific harms, including the promotion of suicide, eating disorders, substance abuse, and certain unlawful products.
• Requiring companies set youth user's accounts to the strongest, most privacy-protective settings by default.
• Allowing kids and teens to turn off data-driven recommendation algorithms.
• Creating tools to help parents track screen time and spending; creating emergency reporting systems that have specific time frames for response.
• Mandating large social media companies to perform annual audits to assess risks to minors and allow qualified researchers access to platform data.

Increased content moderation through policies like KOSA could have mitigated the risks to my son's mental health. Congress must take long overdue action now to protect our children from online harms.



Parent Name: Sondra Worthley
Child Name: Xander Worthley
Rensselaer Falls, NY
My child was harmed on YouTube in 2020 and 2021

Social media has destroyed my family because YouTube allowed a video to stay uploaded because no one was harmed in the video that my child found.

Xander found this video and thought it was safe to try. My son accidentally died March 21, 2022, 3 months before he was to become a teenager. Xander will never get the chance to be a teenager, a freshman in highschool or a licensed driver and so much more because of the harms on social media. I would like lawmakers to hold big tech companies accountable for letting their tech be dangerous and harmful for children and young adults.

I would like big tech companies to make their tech safer for young children and young adults. I would like Big tech companies to stop all the harmful content from being allowed to be uploaded and KEPT up just because no one was harmed. I would like for big tech companies to think of the children that could be harmed instead of how much money it would make them.



Parent Name: Erin Popolo
Child Name: Emily Michaela Murillo
Kendall Park, NJ
My child was harmed on Instagram, Snapchat, and Zoom in 2022

My 17 year old daughter Emily died by suicide after being ruthlessly bullied online through Instagram and Snapchat. During her funeral, her bullies logged into her zoomed funeral (due to covid restrictions) and continued to bully her and say and post atrocious things to the point that we had to stop her funeral. As if our family wasn't devastated enough. The very thing that helped kill my daughter was used to continue to obliterate our lives as well. I would like to see parental controls put in place for these social media

sites. I would like to see the social media company's held accountable for the content and usage of their product and their platforms.



Parent Name: Jason Scheffer
Child Name: Waylon Scheffer
Huson, MT
My child was harmed on WhatsApp in 2020

Waylon interacted with what he believed to be a young girl on WhatsApp in the early morning hours of December 14, 2022. They swapped explicit photos.This alleged ""girl"" was actually 3 sextortionists from Nigeria. They bribed him all night long telling him his life was over if he didn't cooperate and pay them. They were sending him screenshots

of all his social media contacts and threatened him that they'd send them the nude photo of him if he didn't pay up. They convinced him this would make the newspaper.

That morning after I left for work and shortly after that, Waylon drove up in the mountains and took his own life. Meta allowed these phony accounts to attack my son and it took about 10 hours and 30 minutes before he was no longer with us.

Waylon Scheffer 11-01-2006 to 12-14-2022. Lawmakers need to make the social media companies accountable for these horrible things. We've lost a lot of kids due to this crime all while big tech is fattening their pockets.



Parent Name: Jeff Van Lith
Child Name: Ethan Burke Van Lith
Sammamish, WA
My child was harmed on TikTok, YouTube, and Fortnite In 2019

My 13 yr old son died from the "Choking Game" an avoidable death if/when the general public are finally made aware of it. If/ when these kinds of "challenges " are removed from social media on these platforms that are allowing young minds that aren't fully developed yet to have access to this content. Clearly, there are horrible horrible people out there that put these kinds of things on social media, and we can't control that. But we can control what age kids are able to access this and we can control and hold people accountable for putting that on their site to begin with. My son's death was avoidable, and it's time to do something about it. Hold people accountable, for what is allowed on their website, and control with age limits what kids can access.



Parent Name: Todd Minor
Child Name: Matthew E. Minor
Accokeek, MD
My child was harmed on TikTok and YouTube in 2020

Our son, Mathew Emmanuel Minor (Mat), was a very loving & compassionate child. He was a big hugger. Mathew had a very charismatic & caring personality and had a wonderful smile that would light up the room. Mathew was active in martial arts, football, and basketball. He cherished his time at our family gatherings, especially those at the family farm in Tappahannock, Virginia.

On the evening of March 7th, our world was devastated and forever changed by our son TJ pleading to us to come upstairs and that something terrible had happened to

Mathew. Although my wife Mia was starting new treatments for multiple sclerosis, and I was recovering from a recent surgery to remove cancer that subsequently resulted in a leg injury, rendering me unable to walk, the adrenaline kicked in, and I was immediately running upstairs without my walker to check on Mat. My military training kicked in, and I started assessing the situation. Mat had something wrapped around his neck. Why was that there? Find out later, I told myself. I removed the cord, and I began CPR. Mia called 911. While doing this, we asked the almighty why this was happening and to take us instead; he was only 12. I continued to perform CPR until the EMTs showed up. Around 2:30 am at the hospital, the doctor came in and told us that Mathew had died. It was unbelievable and shocking; we had just eaten dinner. The Kids Online Safety Act represents a critical step in shifting the burden of online safety from resting solely on families to being shared with the digital platforms themselves from the moment they design their products. Tech companies would have to design any app or website intended for young users with kids' best interests in mind — including avoiding manipulative design features and ensuring that children's privacy settings are set at a high level by default.

As we've connected with parents across the country, it's become clear that what happened to Matthew was far from an isolated incident. Instead, it was part of a disturbing pattern of kids' deaths linked to mental and behavioral health struggles that the U.S. Surgeon General has called a national crisis fueled in part by social media. In the name of profit, these companies maximize our kids' engagement with online platforms at any cost. We can't count on these tech companies to police themselves; that hasn't worked.

We need legislative action to force social media companies and other digital service providers to prioritize the well-being of their youngest users. That means corporate profits cannot override mitigating foreseeable risks of harm to kids — like avoiding algorithms that push unsolicited videos about dangerous challenges. Until something changes, these companies will keep maximizing our kids' engagement at any cost, all in the name of profit.

There's no time to waste, which is why we're encouraged to see growing momentum across the country; with progress on the national level with the Kids Online Safety Act, that could prevent other families from enduring the kind of pain ours has faced. "



Parent Name: Kristin Bride
Child Name: Carson Bride
Mesa, AZ
My child was harmed on Snapchat in 2020

Social media has impacted our family in the most horrific way possible. In June 2020, we lost our 16-year-old son Carson to suicide after he was viciously cyberbullied by his high school peers who were using Snapchat's anonymous apps Yolo and LMK. In the weeks leading to Carson's death, he received hundreds of humiliating, harassing and sexually explicit messages. The last search on Carson's phone before he ended his life was for hacks to find out who was messaging him.

When we learned that Carson had been cyberbullied over Yolo, we reached out to the company on 4 occasions. We let Yolo know what happened to our son and asked them to follow their own company policies that stated they would monitor for cyberbullying, reveal the names of those who do so, and ban them from the app. We never received a response from Yolo.

After filing a National Class Action Lawsuit in May 2021 against, Snap Inc., Yolo and LMK, Snap Inc. immediately suspended Yolo and LMK from Snapchat. In March 2022, Snap Inc. announced that they were banning all anonymous apps and features from Snapchat because they could not mitigate the risks at an acceptable level. It should not take grieving parents filing lawsuits on behalf of their dead children to motivate social media companies to make changes to keep children safer online. These companies have repeatedly demonstrated that they will NOT self-regulate and make platform safety a priority for hundreds of millions of young users.

Social media platforms know that their products are harmful for young users. For example, anonymous apps have a long history of leading to cyberbullying and suicides when marketed to teens. Clearly, these companies do not prioritize safety when launching a new product like other American industries are required to do.

For this reason, we need the Kids Online Safety Act (KOSA) with its 49 bipartisan senate co-sponsors to be brought to the Senate floor for a vote immediately! This federal legislation is long overdue and will FINALLY require social media companies to have a "Duty of Care" when developing new products to ensure that they are preventing and mitigating harms like cyberbullying and suicide. KOSA will also require that these companies respond to concerned parents and young users when they are reporting an online harm. If the Kids Online Safety Act was in place in 2020, we would still have Carson, our beautiful son who we deeply miss every day. Without federal legislation, Carson and the many other children who have been harmed or have died from these platforms will sadly become just an ever-growing list of Social Media Platforms' "collateral damage.""



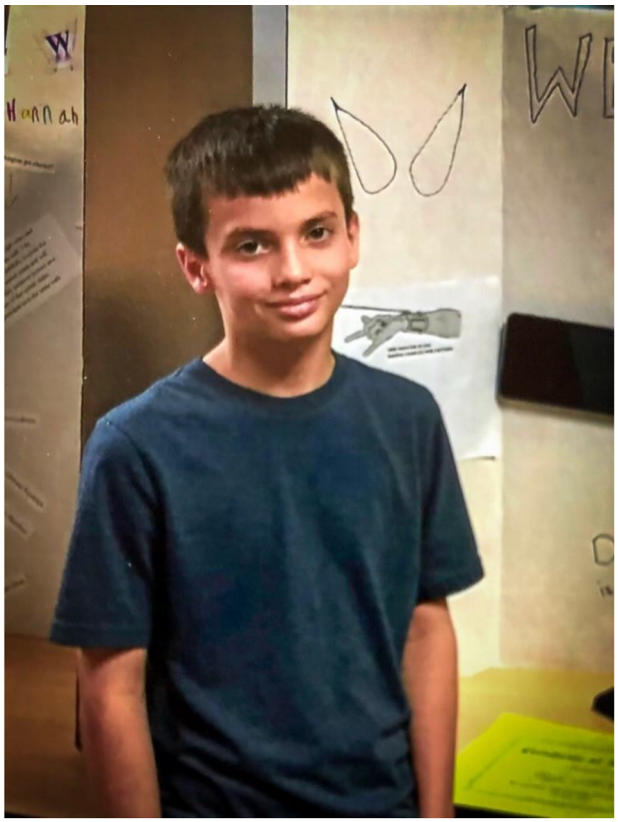
Parent Name: Jennifer Markus
Child Name: Braden Patrick Markus
Lewis Center, OH
My child was harmed on Instagram and Google Hangouts in 2018

My name is Jennifer Markus. On October 17th, 2021 our son Braden took his own life. Ten grueling months after his death, we discovered he was the victim of sextortion. We could not access his phone due to privacy restrictions with apple, which prevented us from taking action to prevent his tragic fate.

Braden was friended on Instagram by someone he thought was an 18-year-old girl from a rival high school, but this person was not whom they appeared to be. Within minutes of exchanging messages, the conversation took a disturbing turn, and within 27

minutes, Braden took his own life. He was 15 years old. Sextortion is the fastest-growing cybercrime in the United States, and it is defined as the practice of extorting money or sexual favors from someone by threatening to reveal evidence of their sexual activity. Our family had never heard of sextortion before Braden's death, and if we had been aware of it, we could have warned our son about its dangers, potentially saving his life. It is essential to take action to protect our youth and their future in the ever-growing digital world. Children are using harmful apps beyond our comprehension, and the Social Media Parental Notification Act, championed by Lt. Governor Jon Husted, would prevent kids under 16 from signing up for social media apps without parental consent and hold companies accountable.

There should also be laws regarding unlocking phones when dealing with minors. Braden was 15 at the time of his death, it took us 10 months to get his phone unlocked, giving his perpetrators 10 more months to continue extorting other innocent kids, and 10 more months to run from law enforcement. We can only hope these innocent children eventually get the justice they deserve!"



Parent Name: Annie McGrath
Child Name: Griffin McGrath
Madison, WI
My child was harmed on YouTube in December 2021

My name is Annie McGrath. Five years ago my only son, Griffin Glen McGrath ("Bubba" to us), asked me to give him an easy math question, then grabbed a snack and went to his room to study, chat, and play games with his school friends online. He accepted a dare from a classmate to try a viral trend called "the pass-out challenge" and it went horribly wrong. Never in a million years did I think that he would never come back

downstairs. Now our hearts are permanently shattered, and we grieve this immeasurable loss every day.

Bubba was only 13 years old. He was an extraordinary and wicked smart child. He had placed third in the National Science Bowl competition just two weeks before he passed. He could solve a Rubik's cube in 8.7 seconds. He was engineering savvy and spent time creating inventions such as a spiderman web-shooter that had the capacity to hang our rocking chair from the ceiling. He crafted his own faster, smoother, more durable Rubik's cube that he sold online to other "cubers". He was a talented drummer and baseball player and took after his father. Most of all he was a kind-hearted soul and touched everyone he met with his brilliance, genuineness, and quick wit. He was a true blessing and as his mother, I did not see any danger of him frequenting YouTube to look at other inventions, engineering tutorials, and cubing techniques. Little did I know of the "invisible danger" that is right in our children's rooms as they study at night.

"The pass-out challenge" has taken thousands of children's lives and is widely available on social media for children to see. A recent Bloomberg article verified that in the last 18 months "the pass-out challenge" has caused the deaths of 15 kids ages 12 and younger and at least 5 deaths of kids ages 13 and 14. This is just the tip of the iceberg, and this preventable issue has greatly impacted Wisconsin families - to which I can personally attest. I was interviewed by a local news outlet regarding "the pass-out challenge" and shortly after the clip aired, I was contacted by 5 other Wisconsin parents who shared that they had experienced the same tragedy.

Not a single social media regulation has been passed in 25 years. Despite the innovation and growth of social media and technology over that time, protective measures and legislation are static. It is inexcusable and unconscionable, period. KOSA currently has 49 co-sponsors and bipartisan agreement. It is long overdue and places a "duty of care" on the tech giants that requires them to act in the best interest of our children and protect their safety and well-being. It also names the FTC to regulate social media going forward, forcing the tech moguls to share their research and allowing an independent audit every two years to ensure that they are compliant. It also thankfully includes a provision that ensures that a separate department would be created to record and respond to reports of dangerous challenges/harms. These new regulations are guaranteed to save children's lives.



Parent Name: Rose Bronstein
Child Name: Nate Bronstein
Chicago, IL
My child was harmed on Snapchat in 2018

My son Nate Bronstein, forever 15, was viciously cyberbullied over Snapchat by classmates at the Latin School of Chicago. Through Snapchat, Nate was threatened

with physical and deadly harm. The cyberbullying included a Snapchat message sent to Nate by a classmate directing him to "go kill himself" . A separate Latin School of Chicago classmate created a Snapchat message intended to harass and humiliate Nate. The Snapchat message included threats, insults and derogatory messaging.

Once the Snapchat message was posted by the creator, screenshots were taken of the Snap by at least 6 other classmates. Every time a classmate picked up the Snap message more insults and threats were added before each student reposted the Snap to their story. One student added an emoji to the Snapchat message that implied to "smoke Nate's ashes". Another student claimed that he reposted the Snapchat to 200 of his friends.

We know the Snapchat went viral and reached hundreds upon hundreds of students even outside of the Latin School of Chicago community.

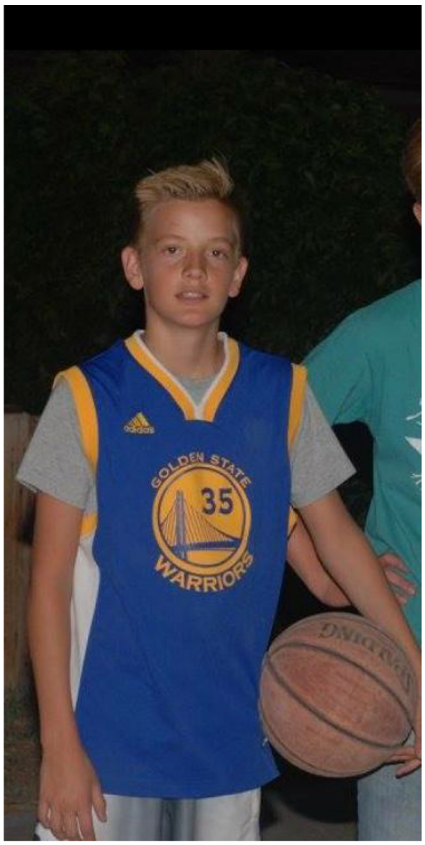
Exactly one month later, Nate Bronstein died by suicide at age 15. Lawmakers need to move quickly and pass a myriad of legislation to protect our nation's children from these horrific harms - which are all preventable. Lawmakers need to get KOSA, the Kids Online Safety Act passed which places a duty of care on social media platforms. Social media platforms have created a dangerous and sometimes deadly environment for tweens and teens through their algorithms and addictive features and yet many of these platforms market themselves to tweens and teens.

These platforms need to be held liable.

The more time passes, more children are harmed through social media platforms and in many cases, the harms are resulting in death.

We need to be unafraid to ask whether the harms simply outweigh the benefits for young teens when it comes to social media. Do 13 and 14 year olds really need to be able to send self-deleting messages to each other? No other product would be able to be sold in a store in America today with this many children having died because of the product.

Social media platforms need to stop fighting sensible common sense regulation meant to save children's lives.



Parent Name: Sage Roberts
Child Name: Keston Brooks Roundy
West Valley City, UT
My child was harmed on TikTok, Instagram, Facebook, SnapChat, and YouTube in 2020

My beautiful boy Keston did not have access to social media but his friends did. They would time themselves and share how long they would last before the so-called “tap out” would take place. On Feb 8th 2018 my daughter found Keston with a cord around

his neck and a knife laying next to his hand. He had tied the cord wrong. It wasn't a slip knot and it wasn't a game. It took my son's life, and so many others.

My world was forever changed and my family. We have managed to piece ourselves back together but we are forever broken. Time cannot heal what happened that day. It doesn't get better. We only learn to maneuver through the hell a bit better.
Please do whatever it takes to prevent these senseless challenges from getting in the hands of our youth. Help us advocate and educate and keep our kids safe. NO ONE DESERVES THE HELL WE LIVE DAILY!

Take this seriously. This needs to be in schools and we need to talk about it just like we do gun violence, drugs, teenage pregnancy, domestic violence etc! The kids already know about it . Education and awareness is vital to the safety of our children! "



Parent Name: Amy Neville
Child Name: Alexander Neville
San Tan Valley, AZ
My child was harmed on Snapchat in 2021

One sunny day in June 2020, I was preparing to take my fourteen year old son, Alexander, to the orthodontist. I went to his room to wake him. There he laid, looking like he was asleep on his bean bag chair, except he wasn't sleeping. Alex was dead. His father tried CPR and paramedics tried naloxone, but it was too late. Alex had taken a single counterfeit Oxycontin that had enough fentanyl to kill four adults.

It had barely been fourteen years. That amazing child, who could do anything he set his mind to, was gone. Prior to his end, Alex skated, scouted, and experimented. Alex was curious about everything and would master subjects. At points in his life, he was an Egyptologist, a Civil War historian, a Pokemon encyclopedia, and so many more roles. By the time we learned he had been using cannabis, he had mastered Snapchat.

Through this app, Alex was able overcome the natural limits that keep most children from trying the hardest drugs. The natural limits that exist for his generation and others include a supportive family, a good school, a good community, and all the other safeguards that we could provide.

Social media transcends these natural limits though. With Snapchat, Alex's normal circle of friends expanded further and began intersecting with abnormal circles. It was on Snapchat that Alex was able to visit with dealers and other users. It was on Snapchat that he set up a deal to get pills. It was on Snapchat that he made plans to have the dealer drive up to our house so that Alex could sneak out for a couple minutes one night and get anything he wanted.

On Snapchat, messages are truly ephemeral and are deleted as soon as they are viewed. Taboo speech, backstabbing, narcissism, etc can all be conveyed through the app with little concern for consequences because the history vanishes after it's written.

Drug dealers, child abusers, and other criminal users appreciate these privacy features too. For them, Snapchat is a portal to other markets that were previously inaccessible and they readily take advantage.

Snap is free of liability for this behavior on their platform because our laws were written to protect them. The only part Snap has ever been accountable for is to answer law enforcement subpoenas. Yet, according to police, sheriffs, and federal agents that I've spoken with, many times Snap is very slow to respond to court orders if they respond at all.

In reaction, Snap posted public safety announcements on their platform with minimal reach and minimal efficacy. More importantly, through a network of lobbyists, lawyers, and $100 billion portfolio, Snap has assembled a well-oiled lobbying machine and attempts to avoid any public policy changes that might assign some responsibility for the deaths that their platform facilitates.

To paraphrase Francis Haugen: for companies like Snap, lobbying and litigation are simply the cost of doing business.

As of today, this is the current map of known deaths where a drug dealer using Snapchat has killed a customer with a fake pill. This includes all counterfeit pills with the majority being fentanyl. Again, this is for *deaths* attributed to pills and Snapchat. It does not include the millions who survive and the tens of thousands more who will die before the year is out.

I ask you, if an airplane company killed this many people in just a few years, would their planes be flying today? Or would we have grounded them until we were sure they were safe? If a car company's vehicles killed drivers and passengers in this many states, would there not be a call for legal and financial accountability? An investigation that resulted in changed business practices? We wouldn't ban planes or cars forever, but we'd make sure before we put people back in them that they were fixed.

Today, Snap is at an inflection point.

Congress can force them to accept responsibility and do better. I have some ideas about that, and I hope during this session we get to explore all of them.



Parent Name: Tammy Rodriguez
Child Name: Selena Rodriguez
Enfield, CT
My child was harmed on TikTok, Instagram, and Snapchat in 2020

On July 21, 2021 we lost our ray of sunshine. Selena Rodriguez was 11 years old when she took her own life in our own home in Connecticut.

That was the final step. She felt she needed to make society like her… Social media ruined Selena's life. It did just what it was intended to do. Get her attention, get her addicted, exploit her for their gain, and now she's gone. This is the social media CEOs earned their income based on. They have stated they would never allow their children on social media. But they sure don't have a problem addicting ours.

We will never know what Selena was going to grow up to be. But I can tell you it was going to be something big. Selena did nothing small! Her life was about helping those

who felt small. Yet ended up being one of those who felt small in the end.It’s time that the lawmakers stop being scared of these” businessman”.

They need to start thinking about our future our children. Because of the adults can’t step up and protect our children, who will? And the reason of these lawmakers are in office are because of us voters.It’s time to think about who will be in office if we continue on the path we’re going now… those CEOs’ kids will be. "



Parent Name: Stephen Carnes
Child Name: Eryn Carnes
Cave Spring, GA
My child was harmed on Facebook, Discord, https://dinostorm.com/en/

Through Dinostorm Eryn was contacted by an individual who convinced her to send nude pictures of herself to them. When the individual noticed blemishes on Eryn's skin they convinced her to kill herself. After my Wife posted about Eryn's death someone gained access to her facebook account and scrubbed all information that was there except one picture. Humbly, I ask that laws be enacted that ensure that individuals who

use technology to harm children be severely punished and if ever released they be heavily restricted from using social media. Please realize that children are precious and deserve to be protected as a national treasure.

Parent Name: Anonymous
Child Name: Anonymous
My child was harmed on TikTok, Instagram, and SnapChat in 2021

My daughter Camryn, age 17, was diagnosed with Anorexia with Bulimic tendencies four years ago and words can't describe how much this disease has harmed my family. Camryn has had a cell phone for the past six years. She has also been in and out of lengthy residential treatment for her eating disorder four times. Nothing can cure this disease. Every time there seems to be a glimmer of hope, the cycle of pain restarts. When I saw Frances Haugen, the Facebook whistleblower, be interviewed on the tv show 60 minutes in October, 2021, my heart stopped. I knew then and there that Instagram was the cause of Camryn's illness. I immediately snatched Camryn's phone from her and saw the unimaginable... a barrage of "pro-Ana" content crowding all of her social media accounts: weight loss advertisements, weight loss challenges, weight-shaming videos, images of near-starving/ skeletal teenagers in various states of undress complaining about how fat they are, etc., etc. These social media companies must be held accountable. They need to be reigned in. Please help.Pro-Ana content must be banned entirely from all social media. The social companies must actively search for such content, delete it immediately, and delete the accounts of the people and companies that propagate it.

Parent Name: Anonymous
Child Name: Anonymous
My child was harmed on TikTok, Instagram, SnapChat, Discord, and Roblox in 2021

In March of 2021, while we were supposed to be safe at home, we discovered that our child was being sexually exploited online via Instagram at the age of 12! We discovered the exploitation was happening when she got a call from Instagram. The predators found her on Instagram, reached out pretending to be a kid her age looking for a "friend", and convinced her to share explicit photos and videos of herself. After a month or so of receiving and sending explicit messages with him, he bullied/ blackmailed her into giving him the login information for her Instagram account so he could use her

profile and photos to lure and exploit other kids. She was scared. She knew this was wrong but didn't know what to do. She had nowhere to turn, she couldn't trust anyone. They told her they knew where she lived and went to school and if she told anyone what was happening, they would kill her family and post all the explicit pictures and videos she sent to them to her friends at school. My husband and I were distraught, she was humiliated and terrified! She was afraid to go to school or leave the house. She had been having stomach aches, and headaches and would think of any excuse to miss school.

When we found out about the exploitation we didn't know where to turn. I googled "how to report child exploitation" and got info for the National Center for Missing & Exploited Children (NCMEC)'s website at www.cybertipline.com or call 1-800-843-5678. I went online and called to report this incident. I was informed that they would reach out to our local police who would follow up on the matter. We spoke to her therapist about this as well to try and get her help right away. Luckily she was already seeing a Therapist for ADHD and Depression. They reported this incident to DCYF as they are mandated reporters. We met the local PD and filed a report, they asked for her phone and login info to use as evidence, to try to get access to her exploiters and after several months there was nothing they could do. The IP addresses were from international areas such as India and they have no jurisdiction internationally. So those predators are still out there! We have multiple bills lined up in favor of online child protection so what's taking so long?

It's been almost 3 years and she is still currently dealing with a lot of that trauma and has even had thoughts of committing suicide due to the overwhelming stress and fear. We want to make sure this can't happen to anyone else, but still protect her from the harms of the public by keeping her identity private. We are lucky we still have our daughter, but I can't say the same for so many others who have lost their children due to the traumas of online exploitation.

We need legislation to pass any of the multiple bills that have been presented to provide online safety for our kids. Something has to be done. If we do nothing then our legislators are guilty of being complicit in these heinous crimes. Legislators must hold Tech companies accountable for providing safe spaces for our kids, monitoring CSAM, and promptly reporting any explicit materials.

Social media companies, I urge you to take action NOW and provide all necessary measures to protect our children on your platforms! Give them a safe space to learn and grow, not a place to be exposed to inappropriate content and fall victim to horrendous acts of sexual abuse. Monitor for child sexual abuse materials (CSAM), provide safety

by design, and provide parents with monitoring technologies and reporting resources. You created these platforms and therefore must provide the changes that are necessary to keep our kids safe from online harms. Help us help them!"

Submission for the Record — Senate Judiciary Hearing, Big Tech and the Online Child Sexual Exploitation Crisis
Sydnie Collins
January 29, 2024

Chair Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

I am about to tell you a story of how I turned my greatest self-destructor into my greatest reason to live. Before I could even spell the word "identity," it became my God-ordained duty to become the quintessential example of the perfect daughter, student, friend, and human. Do you need someone to fulfill a duty and exceed your expectations every time? I got you! Do you want someone to stand in as an unfaltering supporting character for every bump in the road? I'm your girl! I carried these responsibilities with no problem. However, I'd let the slack eat me up inside. I took the saying, "**You** are your biggest critic," seriously, but not how it should be perceived. I challenged myself every day. I wrote notes to myself that proved my bullies right. If there was any opportunity to denounce my name and thrash all the good, potential, and space for growth left inside of me, I'd take it. I wore the word "failure" like a name tag. Eventually, the emotional baggage and anxiety from my verbal abuse caught up with me, but I was supposed to be the strong, self-sufficient, and unmoving pillar in everyone else's world. And if that pillar ever weakened, far be it for me to ask for help.

I was introduced to Instagram at a young age and begged my parents for the app before my wish was finally granted on my 10th birthday. I was in the 4th grade. Almost immediately, I discovered social media's negative impacts on me, but I wasn't sure how to handle it yet. By the time I turned 12, Snapchat became my holy grail. I'm not sure why I wanted it. All I knew then was that my friends had it, so I wanted it too. I didn't fully understand then that people took their own lives because of what people would do to them through a glass screen. I was confused. Why, suddenly, does the thing I wanted so badly make me feel worse about myself than I already did? Within those years, I adopted thoughts of darkness and welcomed them with open arms, not taking the time to find the culprit: social media. With every swipe, double-tap, and snap, I became desensitized to my constant feeling of self-hatred; instead, it became my new normal. It begs the question, how can something so unphysical cause the most pain? To dig deeper, how bad and how often does something have to hurt before you can't feel it anymore?

In the past minute and a half of my speaking, you all have learned that I didn't have the most positive self-image. On the surface, I was intelligent, creative, and "mature for my age." But internally, there was a black hole that sucked every ounce of self-love I had that eventually became patched with insecurity, which was then buried with the burden of upholding other's image of me. I only liked myself because people validated me. I only liked myself because, after every simple Instagram post, I would receive a few more likes than I did before. I slowly became

obsessed with the numbers– comparing them to my friends and especially my older sister, whom I used to idolize. I'd ask myself why some had more followers or likes than I did. How do they have so many friends on Snapchat? How did they get their "snapscore" so high?
The obsession and curiosity grew into an addiction in my early high school years. I had been fighting my suicidal thoughts for years, arguing with the thoughts in my mind if living another day was worth it. I became so lost inside that "identity" became foreign to me. And as morbid as it sounds, you can't wish death on someone who already wants to die; you're just adding fuel to the fire. I had convinced myself that if others hated me as much as I did, what was the point? As I matured, the arrow of my questions pointed less at them and turned toward me. I stopped blaming my agony on others and softened my gaze when I looked at myself in the mirror. In October 2020, I finally asked myself, "Why do *I* care so much?" It didn't occur to me until then that I was giving social media the power to control my emotions. Instagram became a world of fear and distrust for me during the pandemic. I felt increasingly alone and lost whenever I closed out of Snapchat. And don't even get me started on TikTok. I inherited imposter syndrome and a fear of missing out through what I saw on social media. I grew into a routine that hurt me. And every time I turned on my phone, I was eager to see what would hurt me next.

I carried my cross and covered my insecurities with a smile all this time. A smile that hid years of suicidal thoughts that grew more persuasive at each shuttering breath drawn from my earthshattering but silent cries for help erupted from my absolute disdain for myself as a 12-year-old girl. I wanted to offer a gift that I swore everyone would appreciate: my silence, invisibility, and obedience. I didn't know then that the best gift I could ever give anyone is to exist loudly, shine brightly, and live passionately. Oh, the liberation I felt when I gained the strength to delete Snapchat for good. While fighting with my own negative thoughts in my dark days, I was also fighting with God. He was telling me to stay when all I wanted was to go, but He kept me here because He knew my purpose was greater than my likes and followers on Instagram. I learned to appreciate my own existence outside of my life on social media so others could, too. I had to understand that I was made to live instead of hide.

During the pandemic, the isolation, fear, and worry that I, and possibly others, were experiencing moved me to do something about it. At 15, I was still learning how to openly talk about how I felt, and what better way to do that than to open a platform so youth my age could do the same. I created a safe haven for teens and youth everywhere to speak on what is most important to them, and I named it Perfect Timing Podcast, through conversations about entrepreneurship, leadership, human rights, and, most importantly, mental health. For the past 3 years, interviewing the guests on my podcast became therapeutic as I learned after each episode that I wasn't alone. This international podcast I created is proof of how I decided to forgive myself for every negative thought, word, and shadow I cast upon myself, as it wasn't a result of anything but being perfectly human. I turned what caused me the most anxiety, pain, and fear into my reason to keep

going. It became my “why”.

So, I have a new duty: “gnothi seauton,” which is Greek for “know thyself.” No matter how many people I reach, the praise I receive, and the negativity I face through a glass screen, I will continue to know myself and why I decided to fight the loneliness epidemic teens are facing daily. It’s crazy to think the voice that I used to hate became my own remedy because it saved my life. My “why” is to continue to be a voice for those that lost theirs and be the reason that more find it.

Sydnie Collins

Submission for the Record — Senate Judiciary Hearing, Big Tech and the Online Child Sexual Exploitation Crisis
Alix Fraser
January 29, 2024

Chair Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

My name is Alix Fraser and I'm the director of the Council for Responsible Social Media at Issue one. I want to sincerely thank you for your attention to the online safety of children and teens. I am submitting this note for the record alongside other parents and young people who have been personally impacted by social media. These are the real consequences and manifestations of social media. This is why we formed the Council; we see a technology that is built around addiction and profit with little care for the ramifications in their wake. Moms like Kristin Bride and Mary Rodee, as well as powerful Gen Z leaders sharing their stories, reliving the worst days of their lives over and over again because they feel they have to advocate, so others don't have to endure what they have is a true inspiration. It gives me motivation every day to do this work.

But sadly, these are not isolated incidents. One third of Americans know someone who has been harmed by social media. If you're a mother that rises to just under 50%, and if you are a Gen Z individual, or an adult under 30, it rises to a whopping 64%. At least one of the three of you knows someone who's been harmed by social media. And if you're Gen Z, likely two of the three of you knows someone who's been harmed.

We know that social media has changed everything about the lives of young people, ranging from how it consumes their social lives, how they communicate with the world and each other, how they go to school, what's happened in the classroom, how it has disrupted faith communities, how it has impacted and deeply embedded itself in our homes in profound ways. It has impacts we don't even fully understand today. There truly are no spaces that have not been touched by social media for young people. And I know myself, all of us have something deeply personal in this fight. I am, very fortunate to be the father of two young daughters, beautiful young daughters. And, like many of you other parents, they are the lights of my life, and I am absolutely terrified about raising them in this era. They are both younger than three years old, and I am terrified about what the world will look like. Unless we continue to have this fight and make real progress. But I know that I'm not alone in this.

Beyond these harrowing personal stories we've submitted for the record, and all of the individual trauma, we know that is happening, there are broader ramifications for society and for democracy. Issue One, the organization I work for, came to this issue of social media reform, looking at this through the lens of how the information environment impacts our democracy and

the ways we can see the manifestations of that. Right now, wars are raging across the world where no one can tell fact from fiction online, particularly young people. But the deeper we dug into the impacts of social media on youth, the more we realized this was a profound democracy problem in and of itself. Because what does the next generation look like if things continue in this way?

What does the next generation of our democracy look like if young people are addicted, polarized, politically apathetic, depressed, and can't tell what is real and what is false online, and, then, manifest that in real life. It is for that reason that we came together to build the Council for Responsible Social Media, which is a bipartisan group of leaders from many different sectors who've come together across the aisle, across the globe, to try to really make an impact and change things to build a tech and social media world that is accountable, that is transparent, and that is safe for all, particularly for our young people.

And with that, I just want to say thank you for your attention to this crucial issue. The hearing you are convening with the CEOs of these platforms is an important step in demanding accountability, transparency, and responsibility from these companies. But we have heard their promises before. For parents like Mary and Kristin, for young people like Arielle and Trisha, and for the future of our democracy, Congress must finally act to safeguard kids online.

Alix Fraser

Director of the Council for Responsible Social Media at Issue One

Submission for the Record — Senate Judiciary Hearing, Big Tech and the Online Child Sexual Exploitation Crisis
Arielle Geismar
January 29, 2024

Chair Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

My name is Arielle Geismar. I am a digital and mental wellness advocate, and tech policy strategist. I have been organizing around issues that matter to young people since 16, encompassing areas in mental health, LGBTQ+ rights, women's rights, technology ethics, and more. As an advocate, I work with government agencies, startups, nonprofits, and other stakeholders to advise on youth, ethics, and tech interactions. As a digital metal health educator, she hosts daily self care check-ins on my social media. She has been honored to be one of Teen Vogue' 21 Under 21, and is a core member at Design It For Us, a first-of-its-kind youth coalition to achieve key policy reforms around safer online platforms.

We're here today to create change. To take action. Or, in my words, to turn the fish around. You don't get that reference yet, but you will in a minute.

Recently, I penned a forward for a report with Fairplay, documenting how LGBTQ+ youth online face serious harms as a result of manipulative social media platform design. Of many findings, they found LGBTQ+ youth were more likely than their counterparts to lose sleep and track of time due to feeling "stuck" on social media. 24% of Queer young people were recommended dieting and pro-eating disorder content every time they went on social media and 18% received content about drugs or drug sales - every time they went on social media.

My identity is disproportionately targeted and commodified based on intentional decisions. Safeguards are needed, and accountability is crucial. Big Tech companies prey on vulnerabilities they believe would extend our engagement. Their choices correlate to declining mental health and well-being among a community already fighting day in and day out.

Simply put - when queer youth are under attack, we create digital spaces. They are our home, and they are our right. They are lifesaving. The last thing we need is to be unsafe there, too. Protections for queer youth embedded in the design of the digital world would enable us space to grow and thrive.

Mental health advocacy isn't just a passion, it's personal. I lost a close friend to suicide in high school. Frankie was bright, funny, and a major catalyst for my focus in mental health. Mental health and safe online platforms are inextricably linked.

As a college senior now, I'm thinking about how Frankie and I never got to spend this time in our lives together. But, when I think about my experience in student social justice organizing, it feels like she's always been here. So, let me share some of the lessons she taught me with you. The little moments we shared made me who I am today, and moreover have made me the advocate I am today. We can take her lessons in our policy work.

Frankie was the most fashion-forward person I've ever met. She wasn't afraid to be bold, try new patterns, put new pieces together. As we explore policy solutions, be bold. Look for old patterns in what research tells us. Look for new patterns when creating new platforms. Let's try putting new pieces together, even if we're certain they don't go. We may be surprised.

Frankie and I showed up for each other. Supported each other. As we go on through our policy work on creating a more just internet, take care of each other. Check-in with your co-workers. Your friends. Ask - how're you doing today?

Technology ethics organizing can be a bit different than other forms of organizing. At the end of the day, when we're tired and need a break, we become part of the system we're trying to regulate. It's tough to want to just scroll at the end of the day when we know what media consumption does to us. It's not our fault. Don't beat yourself up over it. When it's time to relax, relax. When it's time to fight, fight.

Frankie was incredibly welcoming. As we think about technology, I urge you to welcome young people into your work. Do not create policy around social media about us - without us. Ask us questions, early and often. Let us into your thought process, and take our advice. Young people are a stronghold within this space, and we deserve a seat at the table.

Frankie and I took Marine biology together in high school. One time, we were diagraming a real fish on a tray in front of us, matching anatomy. We were struggling to match the real fish to the diagram, when the reality was much more unsavory than the piece of paper we'd studied on. As we struggled, Frankie took the tray with the fish, and turned it around. Physically. We'd been going about it upside down and didn't realize it, making it much harder for ourselves. One simple act, and our perspective was changed. As policymakers and change makers , we can spend a lot of time looking at a problem, when sometimes the answer is actually in front of us. Sometimes the answer is just working together. Turning the fish gave us a new perspective. And neither of us could have done it alone.

As we go into today, think about the new perspectives you can gain. The conversations you can have, the ideas you can form. It's our responsibility to take care of young people through action. Through policy. Through collaborating and pushing further than big tech tells us to.

Today, and every day, turn the fish. I think you'll see something new. Thank you.

Submission for the Record — Senate Judiciary Hearing, Big Tech and the Online Child Sexual Exploitation Crisis
Trisha Prabhu
January 29, 2024

Chair Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

Good afternoon. My name is Trisha Prabhu, and I am a Gen Z anti-hate activist and the Founder and CEO of ReThink, a global movement with the mission of tackling cyberbullying and redefining internet culture to be more positive. My journey really started back when I was a 10-year-old, growing up in the suburbs of Chicago. At the time, my community was a predominantly white community. And that meant that I knew from a very young age, what it felt like to be excluded, what it felt like to be isolated, what it felt like to be different. That was always very apparent to me. It meant that, as a kid, I faced some bullying and some remarks. And it also meant that when we all got phones, I experienced cyber harassment.

I received mean, disparaging remarks from people I often didn't know or from people who were anonymous about everything from my hair to my cultural background. It was extremely difficult. I was such a confident child. My mom would always say that about me: that I was such a confident child. But when you're cyber harassed, and it's constant, right, it's 24/7, it takes that confidence from you. I suddenly felt like I couldn't depend on anyone. I suddenly felt unsure of myself. I became quieter, more reserved, and of course, I was afraid, so I hid it. I didn't tell my parents. I didn't talk to an adult. I didn't talk to a teacher. I didn't talk to a counselor.

I hid it, and I dealt with it, or at least that's what I told myself. When the abuse finally started to subside, I, like so many victims of cyberbullying, just wanted to forget about it. I didn't want to think about it ever again in my life and I wanted to move on. But it was later, when I was 13, that, one day, I came home from school, and I happened to read online the story of another young woman from a completely different state, who had been cyberbullied for over a year and a half, receiving some of the worst messages I have ever, ever seen: "The world would be a better place without you." "I hope you drink bleach and die." "You are so ugly."

And unfortunately she died by suicide because of that experience. At 13, reading that story, for me, there were two moments of impact: one was, something is seriously wrong here. This is just unacceptable, what is going on. And two, it was a moment of realization for me of, *oh, this isn't just a problem that I'm experiencing*. I had thought for the longest time, *oh, I'm being cyber bullied because there's something wrong with me*. Reading that story was a realization of, no, in fact, online hate is a silent pandemic affecting youth globally. It's something we've been taught to live with, but that we shouldn't have to, because we deserve an internet that is better than this. And so, especially as a young person, I was fired up. I was frustrated.

I felt like there simply weren't enough youth voices in the responsible technology space, that we needed more youth at the table. I was also frustrated by what felt like very reactive, outdated solutions to tackle cyberbullying, like simply encouraging youth to tell a parent about their experience. We know that those types of solutions often don't work, that young people are scared, that they worry that they might get in more trouble if they say something. And so my

vision was, my question was: is there some way that I can redesign the internet, so that the burden is *not* on the victim to act, but instead on the bully to pause and think before they say something offensive?

That is what led to the creation of ReThink, which today builds technology that detects cyberbullying before it's sent and gives users a chance to pause, review, and rethink sending it. Put more simply, we create friction in online communication. The idea is that by giving them that moment to pause, we give youth a chance to think about who it is that they want to be online, we give them a chance to set themselves up to be responsible digital citizens — rather than get sucked into a digital environment that can make it feel like their words don't matter, that can make it feel like there are no consequences when in fact, as we've thus far absolutely seen today, there definitely are very, very real consequences. Building this movement has been so, so incredibly fulfilling. It's been an opportunity for me to take an experience that I wanted to forget and instead turn it into something positive.

It's also illustrated for me the power in investing in common-sense solutions that address cyberbullying. In the case of ReThink, our research, which been celebrated by Google and MIT, among other institutions, finds that 93% of the time, youth aged 13 to 18 that receive a ReThink alert change their mind and decide *not* to say something offensive. It is that simple. And so, this idea, that we often hear from tech companies, that we're stuck with the social media that we have, that we're stuck with the internet that we have, that we can't possibly do better than this because it's all just too complex…that's an idea that I reject. I think it's an idea that a lot of us here today reject. So I'm really, really excited and passionate about what we can do moving forward to find these easy, common sense solutions that can make the internet a more safe place for youth, that can realize the internet that I deserved as a young person and the internet that every young person deserves. Thank you.

Submission for the Record — Senate Judiciary Hearing, Big Tech and the Online Child Sexual Exploitation Crisis
Mary Rodee
January 29, 2024

Chair Durbin, Ranking Member Graham, and Members of the Senate Judiciary Committee:

My name is Mary Rodee, I am a concerned citizen, an advocate for online safe guards, a worried elementary school teacher but mostly as a surviving mother forever haunted by an irreplaceable loss. I connect to the word surviving since there is no escape from this living nightmare. I will never move past it or get over it. I am just surviving the heartache that comes with having a severed love line.

I'm writing to you to share with you the story of my son, Riley. Riley was the life of the party – a joyful, funny kid who loved sports and farming and dreamed of a job in environmental conservation. Riley was brave and took risks, like a lot of teenagers he could be impulsive. His laughter was infectious, his spirit boundless, he could find humor in the ordinary. On March 30, 2021, that spirit was extinguished. At just 15 years 9 months and 27 days old, Riley's life was tragically cut short by a sinister force he encountered on Facebook. A criminal halfway across the world friended Riley, manipulated him into sharing revealing photos, and then, using those pictures, blackmailed him. In a few torturous hours, overwhelmed by fear, uncertainty, and shame, too young to see any way out, Riley chose to end his life.

We have an expansive and diverse family in a rural college town in Northern NY. We live north of the Adirondack mountains just south of Ottawa on a farm that's been in the family for 7 generations. Riley has 11 siblings and step siblings and he is my youngest child. He grew up fast, exposed to a lot of life at an early age, we have been through so much together as a blended family. We have had the talks, we have battled depression and anxiety with our kids. We suffered the loss of friends by suicide, irresponsible teenage behavior and accidents. We have kids with IEP's and kids in counseling. Our kids play the sports and do the musicals. We put in the time, we had the talks, we are available. Riley has a vast support system, including 8 biological aunts and uncles and their spouses, grandparents and cousins most of whom live in our area and some who live next door. Riley is loved by his coaches, teachers and friends and their parents. I know in my broken heart that he knew I would do anything for him and even with this life so full of love, realness and acceptance someone still got to him through Facebook and in only 6 hours, scared him to death.

The investigations by the NYS police and the FBI were able to locate the terrorists who took my baby, but the sad truth? The perpetrators remain unaccountable. The FBI investigators tell us that there is nothing they can do because the US doesn't do business with that country in Africa but yet each and every minute these criminal murderers are allowed to sneak into our country and do

business with our children. And while they are the immediate culprits, it is platforms like Facebook that provided them the tools and access to prey on innocent souls like Riley. And Riley is not alone. Thousands of parents worldwide are waking up to the grim reality of how perilous Meta's products can be for their children.

I am regularly contacted by parents trying to navigate the fallout from their child's sexploitation. They want help and answers and the best I can do is give them reporting tools and tell them to praise their child for coming to them because the disturbing reality is that it is unlikely there will be a resolution.

As a teacher I am scared to see kids placed on computers as part of general education practices while teachers have no tools to protect them. So many are fooled by the fragile facade of social media, lulled into a false feeling of autonomy and security, until one day your whole world comes crashing down. Healthy fears keep kids safe and there is not a fear of what is happening as we allow devices to become an extension of our person, even the youngest children.
Riley's life, his dreams, and his future were not just collateral damage in the world of big tech and business. He is my son. And there are countless Rileys out there, teetering on the edge, vulnerable to the next online predator or scam. Harmful big tech practices killed my child, provisions like the duty of care component of the Kids online safety act could have prevented Riley's death. Protecting children online is not a job for parents or for educators or for police officers. It is impossible for parents and teachers like me to be equipped to fight this. This isn't a sporadic issue; it's a systemic one. We need real safeguards, we need KOSA. Let's demand safety, transparency and accountability. For Riley, for all our children. Thank you.

Statement for the Record of Uldouz Wallace

As a Survivor of Image Based Sexual Abuse I know how urgently it is needed to hold online platforms accountable, we need laws to protect children from getting harmed online.

It's been 10 years since I was exploited by online platforms, I lost family, friends, money, work, reputation, opportunity, relationships the list goes on and on and the whole time online platforms were profiting from my exploitation and everyone else and still there's now laws to hold them accountable, not even children!

Everyone can in an instant become a victim of image based sexual abuse through deepfake, AI.

Currently online platforms have the technology to protect children, but instead they're putting profit over children, women, and men that are being exploited online.

How many more lives of children do we have to lose for online platforms to take action? What if someone you loved was harmed by online platforms?

We need laws that are updating with technology in order to protect the people.

It's time for online platforms to be held accountable.

Dear Chair Durbin, Ranking Member Graham, and Senate Judiciary Committee Members,

The undersigned offer our collective testimony as grieving parents who have experienced the devastating loss of a child due to online harms. Our families have faced the tragic consequences of the inadequate measures taken by Snap Inc., Meta, TikTok, X, and Discord to ensure the safety and wellbeing of children using their products. Our children have fallen victim to drug dealers who openly sell fentanyl-laced counterfeit drugs on their platforms. They have fallen victim to merciless cyberbullying, leaving them so hopeless they choose to take their own lives. They have fallen victim to algorithms feeding content that encourages severe, and at times fatal, eating disorders. They have fallen victim to dangerous challenges that are pushed to their feeds. And, as is the focus of today's hearing, they have fallen victim to the pervasiveness of online sexual exploitation.

Evan Spiegel (CEO of Snap), Mark Zuckerberg (CEO of Meta), Shou Zi Chew (CEO of TikTok), Linda Yaccarino (CEO of X), and Jason Citron (CEO of Discord), can no longer avoid the power of our grief…or our message. We are survivor parents who not only live with the pain of our loss, but with the knowledge that each loss could have been prevented if these CEOs chose our child's safety over their enormous profits.[1] So today, on behalf all survivor parents, family and friends, we demand accountability from Snap, Meta, TikTok, X and Discord.

On November 7, 2023, former Facebook Engineering Director and Consultant, Arturo Bejar, testified before this committee. He spoke of Mark Zuckerberg's blatant disregard for an internal study showing the negative teen experiences on Meta's platforms. Bejar's research found that over 25% of 13-15-year-olds had received unwanted sexual advances on Instagram.

This may be a shocking revelation to some, but not to the signers of this letter. We are well aware of the business model Snap, Meta, TikTok, X and Discord ascribe to: profits over people, which means increased engagement is prioritized over all else. We, and the CEOs being questioned today, are also well aware of the pervasiveness of the sexual exploitation and harassment taking place on these platforms - not only violating the ethical standards that users should expect, but also each site's own publicly stated community guidelines.

As the five CEOs prepare to testify before the Senate Judiciary Committee, we want them to feel our overwhelming pain, anger, and frustration. We want them to understand that today, they not only face the scrutiny of the Committee, but that of all the grieving families whose lives have been forever altered by their actions and inactions; their relentless pursuit of profit at the expense of user wellbeing; their refusal to invest in sufficient content moderation; their manipulation of user behavior through targeted advertising; their harmful algorithm designs; and surveillance practices. All leading to the creation of the ultimate products or manipulation and exploitation of children, turning young users into the collateral damage of their greed. They have completely failed to regulate themselves, and our children and families have paid the price.

At today's hearing, we hope that each CEO will confront our grievances by providing clear, unambiguous, honest answers regarding the steps they will take to prevent the same egregious

harms that our children experienced from occurring to others. We implore each CEO to take this opportunity to demonstrate a genuine commitment to prioritizing the wellbeing of each child, and user, they claim to serve. **The undersigned grieving families deserve no less than this.**

Thank you for your consideration of this testimony.

Sincerely,

Deb Schmill - Parent of Rebecca Mann Schmill (Forever 18)
Stu Schmill - Parent of Rebecca Mann Schmill (Forever 18)
Amy Neville - Parent of Alexander Neville (Forever 14)
Rose Bronstein - Parent of Nate Bronstein (Forever 15)
Rob Bronstein - Parent of Nate Bronstein (Forever 15)
Samuel P. Chapman - Parent of Sammy Chapman (Forever 16)
Dr. Laura Berman - Parent of Sammy Chapman (Forever 16)
Kathy McCarthy - Parent of Jack McCarthy (Forever 19)
James McCarthy - Parent of Jack McCarthy (Forever 19)
Christine McComas - Parent of Grace K. McComas (Forever 15)
Dianne Grossman - Parent of Mallory Rose Grossman (Forever 12)
Seth Grossman - Parent of Mallory Rose Grossman (Forever 12)
Elaina LoAlbo- Parent of Felicia LoAlbo-Melendez (Forever 11)
Kristin Bride - Parent of Carson Bride (Forever 16)
Erin Popolo - Parent of Emily Murillo (Forever 17)
Catherine Williamson - Parent of Sean Williamson (Forever 16)
Daniel Williamson - Parent of Sean Williamson (Forever 16)
Cheryl Brown - Parent of McKenna Elizabeth Brown (Forever 16)
Hunter Brown - Parent of McKenna Elizabeth Brown (Forever 16)
Maurine Molak - Parent of David Molak (Forever 16)
Matt Molak - Parent of David Molak (Forever 16)
Shannon Lee - Parent of Ashlyn Taylor Lee (Forever 16)
Avery Schott - Parent of Annalee Schott (Forever 18)
Lori Schott - Parent of Annalee Schott (Forever 18)
Ian Russell - Parent of Molly Russell (Forever 14)
Deborah LaCroix - Parent of Jonathan Justin Shea (Forever 35)
Joann Bogard - Parent of Mason Bogard (Forever 15)
Steve Bogard - Parent of Mason Bogard (Forever 15)
Juliana Arnold - Parent of Coco (Forever 17)
Jane Clementi - Parent of Tyler Clementi (Forever 18)
Annie McGrath - Parent of Griffin McGrath (Forever 13)
Jaime Puerta - Parent of Daniel Puerta-Johnson (Forever 16)
Judy Rogg - Parent of Erik Robinson (Forever 12)
Sondra Worthley - Parent of Xander Worthley (Forever 12)
Kristie Reilly - Parent of Noah Dennis (Forever 13)
Sabrina Stanesby - Parent of William Stanesby (Forever 12)
David Stanesby - Parent of William Stanesby (Forever 12)
Mary Rodee - Parent of Riley Basford (Forever 15)
Robin Jones - Parent of Zachary Jones (Forever 16)
Michelle Servi - Parent of Jack Servi (Forever 16)

Jeff Van Lith - Parent of Ethan Burke Van Lith (Forever 13)
Sharon Freeman - Parent of Andrew Freeman (Forever 13)
Lon Emerick - Parent of Noah Geiwitz (Forever 17)
Heidi Emerick - Parent of Noah Geiwitz (Forever 17)
Chris Kirkendall - Parent of Joshua Kirkendall (Forever 14)
Jennifer Kirkendall - Parent of Joshua Kirkendall (Forever 14)
Todd Minor - Parent of Matthew E. Minor (Forever 12)
Mia Minor - Parent of Matthew E. Minor (Forever 12)
Christine Lesher - Parent of Sean Lesher Edwards (Forever 17)
Bridgette Norring - Parent of Devin Norring (Forever 19)
Tom Norring - Parent of Devin Norring (Forever 19)
Leah Gallant - Parent of Charlotte Gallant (Forever 19)
Pam Love - Parent of Michael Love (Forever 35)
Andrea Silvano - Parent of Zachary Parsons (Forever 21)
Fred Walters - Parent of Jocelyn Walters (Forever 14)
Rosemarie Maneri - Parent of Shylynn Dixon (Forever 18)
Jennifer May - Parent of Skyler Denney (Forever 24)
Kathy Klingele - Parent of Serquoyah Klingele (Forever 16)
Brenda Reedy - Parent of Jose Ricardo Ellis Reedy (Forever 20)
Robert and Asiah Reedy - Parent of Jose Ricardo Ellis Reedy (Forever 20)
Kimberley Gustavson - Parent of Travis Gustavson (Forever 21)
Robert Connerney - Stepparent of Zachary Parsons (Forever 21)
Tabbatha Urbanski - Parent of Seth Carson (Forever 17)
Erika Shambaugh - Parent of Josha Legassey (Forever 20)
Jeanine Bothell - Parent of Luke Allen Benson (Forever 25)
Michelle Hiser - Parent of Joseph Hiser (Forever 32)
Teresa Reese - Parent of Alden Reese (Forever 15)
Julie Hines - Parent of Tyler Young (Forever 19)
Christina Wagar - Parent of Cassandra Cammarata (Forever 25)
Debra A. Berninger - Parent of Eric R. Berninger (Forever 35)
Lisa Tyler - Parent of Tyler Champagne (Forever 27)
Brenda Hicks - Parent of Dennis Brandon Hicks (Forever 36)
Kim VanderMeeden - Parent of Richard VanderMeeden (Forever 19)
Pauline Stuart - Parent of Tyler Ryan Last (Forever 17)
Chris Didier - Parent of Zach Didier (Forever 17)
Lorie McCormick - Parent of Michael McCormick (Forever 30)
Brian Montgomery - Parent of Walker Montgomery (Forever 16)
Keith Godsey - Parent of Hunter Godsey (Forever 25)
Crystal Godsey - Parent of Hunter Godsey (Forever 25)
Maureen Weston - Parent of Cedric (Forever 20)
Joy Harrison - Parent of Jacob Harrison (Forever 23)
Rebekah Brown - Parent of Cole Brown (Forever 18)
Cindy Cruz-Sarantos - Parent of Dylan Kai Sarantos (Forever 18)
Andy Parker - Parent of Allison (Forever 24)
Dawn Carpenter - Parent of Stone Carpenter (Forever 18)
Tammy Rodriguez - Parent of Selena Rodriguez (Forever 11)
Raquel D'Oyen - Parent of Mitchell (Forever 18)
Michael Drout - Parent of Mitchell (Forever 18)

Julie K. Berndt - Parent of Faith Loren Berndt (Forever 20)
Cathy Haynes - Parent of Jimi Ray Haynes, III (Forever 37)
Tauheedah Washington - Parent of Tajir Williams (Forever 21)
Brian Veal - Parent of Claire Veal (Forever 17)
Kay Schulz - Parent of Nickolas James Schulz (Forever 25)
Rosanne Yackel – Sister of Jessica Lynn Caruso (Forever 36)
Misty Terrigino - Parent of Kaylie Tallant (Forever 17)
Maggie Taylor - Parent of Emily Taylor (Forever 17)
Wendy Plunk - Parent of Zachariah Plunk (Forever 17)
Roy Plunk - Parent of Zachariah Plunk (Forever 17)
Amanda Eubanks-Poso - Parent of Luca Manuel Poso (Forever 13)
Kimberly Osterman - Parent of Max Osterman (Forever 18)
Sabrina Jankowski - Parent of Joseph Jankowski IV (Forever 24)
Aaron Taylor - Parent of Parker Stewart Taylor (Forever 24)
Sarah Taylor - Parent of Parker Stewart Taylor (Forever 24)
Petra Verhoeven-Jordan - Parent of Gian-Luc Jordan (Forever 12)
James McCarthy - Parent of Jack McCarthy (Forever 19)
Despina Prodromidis - Parent of Olivia Green (Forever 15)
Sara Manser - Parent of Ashton Harmon-Manser (Forever 22)
Jennifer Ebert - Parent of Jason Patrick Ebert (Forever 21)
Jim Ebert - Parent of Jason Patrick Ebert (Forever 21)
Tricia - Parent of Chad (Forever 29)
Phyllis Gabbard - Parent of Roman Blake Pelfrey (Forever 26)
Misty Medrano - Parent of Ethan Daniel McCaskill (Forever 10)
Sarah Eager - Parent of Lauren Eager (Forever 18)

1. Raffoul, A., Ward, Z. J., Santoso, M., Kavanaugh, J. R., & Austin, S. B. (2023). Social media platforms generate billions of dollars in revenue from U.S. youth: Findings from a simulated revenue model. *PLOS ONE*, *18*(12), e0295337. https://doi.org/10.1371/journal.pone.0295337

ANFHELP.ORG
Alexander Neville
FOUNDATION
Drug Prevention Education
Social Media Harms Awareness
Fostering positive youth mental
health and resiliency

Beccaschmillfdn.org
Becca Schmill
Foundation
Funds research, sponsors community programming, and advocates for policies that promote and safeguard the emotional wellbeing of adolescents and young adults

Bucketsoverbullying.org
Buckets Over Bullying
Bronstein Family Foundation
In loving memory of Nate Bronstein
To stop cyberbullying of children and teens through education, lawmaking and legal action

Paper One - UK's Approach to End-to-End Encryption to Ensure Public Safety and National Security

Overview:

The UK Government is pro-innovation and pro-privacy, and we continue to champion strong encryption recognising its contribution to privacy and cyber-security. However, the Government's position is that the use of such privacy enhancing technologies should not come at the detriment of the safety of vulnerable citizens, such as children, and national security. Too often the debate on end-to-end encryption (E2EE) has been described as a binary choice between privacy and public safety/national security, implying that you can either have one or the other. The reality is that it is possible to strike a balance that respects both, as well as protecting cyber security and without having a detrimental impact on technological innovation, as long as an absolutist position on each is avoided. One of the principles underpinning the UK Government's approach, however, is that this should be determined by liberal democratic governments who are ultimately responsible for their people's safety.

Preventing the sexual exploitation of children:

Technology and social media enables us to create intricate connections with others within our community, and those around the globe. However, in the UK, our evidence suggests that thousands of offenders are also exploiting the same features and services offered by social media services to discover, target and sexually abuse children at scale. In 2022 alone, technology companies reported a record 32 million referrals of suspected child sexual abuse, which contained over 49.4 million images of child sexual abuse, and these numbers continue to grow year on year.

UK Law Enforcement brought to justice David Wilson, a prolific offender who had targeted over 5,000 children across Facebook and SnapChat with hundreds of his victims being coerced to share intimate imagery and videos. UK law enforcement agencies received over 250,000 messages predominantly from Facebook which displayed Wilson's abhorrent activities, and this crucial evidence supported the investigation and prosecution of Wilson. In 2023, Ishmael Duncan was brought to justice after a period of prolific offences against children online using Snapchat, whereby he contacted approximately 10,000 children across the globe. Mr Duncan coerced and threatened them to into sending intimate imagery, and on one occasion he attempted to coerce a 14-year-old to perform sexual acts on their sibling. Identifying these cases was only possible through the proactive child safety protections companies have in place, that have enabled robust content moderation and identification of abuse.

These cases highlight the need for implementation of robust child safety controls and reporting mechanisms by technology companies to enable them to identify and prevent abuse, and refer any instances of abuse that have occurred to law enforcement agencies to act and bring criminals to justice. However, these child safety protections are currently being eroded through the introduction of end-to-end encryption. The implementation of end-to-end encryption without robust child safety features blinds technology companies to child sexual abuse that continues to proliferate across their platforms, and critically hampers the ability of law enforcement agencies to gain crucial evidence against offenders; evidence which is critical in order to bring them to justice and safeguard victims from abuse. For example, the UK's National Crime Agency estimates that they will lose 92% of Facebook Messenger and 85% of Instagram Direct referrals[1], as a result of Meta's move to implement E2EE with no robust child safety protections on their messaging platforms.

The UK Government has emphasised the importance of a safety by design approach to tackle online harms, and tech companies must take due regard to the risks posed from the design choices, such as E2EE, that they choose to develop and implement. It is crucial that tech companies implement robust child safety measures that maintain the ability to detect the sexual abuse of children, including in an end-to-end encrypted environment.

The UK government-funded Safety Tech Challenge Fund incentivised the tech sector to develop innovative technical solutions that detect child sexual abuse material in end-to-end encrypted environments, whilst respecting user privacy. The fund demonstrated that it technically feasible to implement robust solutions that enable the detection of child sexual abuse within encrypted environments without weakening encryption. Therefore, it is reasonable to expect that companies can innovate further, by investing resources and engineering expertise, in building technological solutions to tackle child sexual abuse material appearing on their encrypted platforms.

Linked to the rise in online child sexual abuse, the UK Government has implemented the Online Safety Act 2023 which brings into force regulatory measures which require companies to take action to minimise the risk of child sexual abuse

[1] NCA response to Meta's rollout of end-to-end-encryption - National Crime Agency

occurring on their platforms. The Act is tech agnostic in nature and applies to user-to-user services irrespective of their design, meaning end-to-end encrypted platforms will also be in scope. Paper Two provides further information on the Act, its purpose and enforcement of the new regulatory regime.

Lawful Access:

The primary responsibility of any state is to ensure the security, safety, and wellbeing of its citizens, including through upholding fundamental rights and freedoms. Critical to this is lawful access, by exception when necessary and proportionate, to communications data held by technology companies. This data is crucial to the investigation, prevention and prosecution of the most serious crime and national security threats: child sexual abuse, terrorism, and drug trafficking to name but three. Just as in the US, our law enforcement agencies depend upon this data to prevent, detect and investigate the most serious crimes effectively. It is vital to ensuring public safety; vital to saving lives.

The content of messages reveals operational details of the activities subjects of interest carry out in order to facilitate terrorism or other serious crimes, such as meeting with conspirators or procuring materials or weapons needed for the crime or attack. Access to content can also provide unique insight into the intention and mind-set of what a subject of interest is planning to do in the future and at what stage of criminality they are. This allows those investigators to take vital action to prevent crimes from taking place.

End-to-end encryption, without lawful access capabilities designed in, can be a threat to public safety and national security. Without such capabilities, this leads to the erosion of lawful access and blinds law enforcement in their ability to investigate the most serious crimes. In 2022 alone E2EE touched on nearly all the 800 live CT Policing investigations in the UK.

E2EE therefore also risks undermining the utility of the UK-US Data Access Agreement (DAA), which allows law enforcement in either country to access critical data held in the other for the purposes described above. The DAA the culmination of bipartisan support and efforts across Congress to enact the CLOUD Act in 2019. The DAA is having a transformative effect on law enforcement and has now been used over ten thousand times, with data provided under the DAA having proved critical in supporting high priority child sexual exploitation investigations. This has directly resulted in the safeguarding of vulnerable children from harm and the arrest of those suspected of committing offences against children.

The UK Government believes that democratically overseen functions critical to public safety must not be undermined by tech companies without consultation. In contrast, the UK Government strikes a balance between privacy and national security through our Investigatory Powers Act 2016 (IPA); a clear legal framework, openly scrutinised and approved by Parliament, that provides our law enforcement and intelligence agencies with the powers they need to investigate crime and defend our national security. At the same time, the IPA ensures that data protection and privacy are at the heart of those powers and that they are subject to strong safeguards and oversight, including independent oversight for the most intrusive forms of data access. This model, of balance, oversight and safeguards, is a far more proportionate and legitimate approach than the blanket loss of data access which some tech firms seek to impose. Further, Parliament are also considering a number of targeted changes to the IPA to ensure it keeps the pace with modern technology, including a new Notification Notice and the need for tech companies to maintain the status quo within a Notice Review period.

Given the importance of these powers and the impact end-to-end encryption is having on their use, like-minded governments and international institutions are increasingly of the view that while the use of encryption is vital, that must not come at the expense of precluding law enforcement from being able to access the content of communications where that is needed to progress their investigations and is subject to robust safeguards and oversight. This includes the 2020 International Statement on E2EE[2], the 2023 G7 Interior Ministers' statement [3], the 2021 Johnson-Atlantic Charter[4], and the 2023 Five Country Ministerial communique[5], all of which declared the necessity to 'maintain tightly-controlled lawful access' through 'working in partnership with technology companies.'

[2] 2020.10.11_International_statement_end-to-end_encryption_and_public_safety_for_publication_final.pdf (publishing.service.gov.uk)
[3] G7 Interior and Security Ministers' Communiqué | Homeland Security (dhs.gov)
[4] UK-US_Joint_Statement_-_June_10.pdf (publishing.service.gov.uk)
[5] Five Country Ministerial Communiqué 2023 (publishing.service.gov.uk)

Paper Two - UK Online Safety Act 2023

Overview:

The Online Safety Act received Royal Assent (became UK law) in October 2023. The Act has the aim of making the UK the safest place in the world to be online. The Act requires in scope services to put in place measures to protect all users from illegal content and protect children from online harm, while empowering adults with more choices over what they see online.

The internet and other new technologies have driven dramatic improvements to our economy and society. The benefits of technological innovation must extend to all citizens; illegal and harmful behaviours and content cannot be allowed to undermine these. Companies will be held to account for tackling a comprehensive set of online harms – with particularly robust action required in relation to child sexual exploitation and abuse and terrorism, due to the exceptionally high potential for harm to users.

The Act therefore has three principle aims:

1. To tackle criminal activity online (including child sexual exploitation and abuse)
2. To protect children online
3. To promote transparency, accountability and freedom of expression

Framework:

The central pillars of the Act are statutory duties, which apply to in-scope user-to-user and search services:

- **Illegal content safety duty**: to put in place safety measures to reduce the risk their services facilitate offending, and to rapidly remove illegal content when it does appear;
- **Child safety duties**: to ensure children are prevented from accessing the most harmful, legal material and have an age-appropriate experience online;
- **Adult user empowerment duties**: tools that adults will be proactively offered to increase their control over the content they see and the users they engage with.
- There is also a standalone duty in the Act requiring providers who publish or place pornographic content on their services to prevent children from accessing that content.

Companies will have to carry out **assessments of the risks** present on their platforms, and be expected to put in place proportionate systems and processes to mitigate those risks and thereby comply with their safety duties. The UK's online safety regulator, Ofcom, will set out specific steps in codes of practice that providers can take to fulfil their new duties.

Child sexual exploitation and abuse (CSEA) specific measures

CSEA is designated as 'priority' illegal content, meaning that companies are subject to the strictest duties for this content: they will have to prevent users from encountering this content. A service must carry out a risk assessment of CSEA on its platform, and take risk-based, proportionate steps to minimise CSEA informed by their risk assessment. Ofcom, will set out steps that companies *can* take to meet their safety duties in its Codes of Practice.

End-to-end encryption and CSEA:

The Online Safety Act is technology neutral and, as such, does not mention specific technologies by name. The Act is thereby designed to be future-proofed, based on safety duties and the responsible and proportionate measures put in place by companies, which are determined by their own service design and risk level, to meet those duties.

A company must assess the risk of illegal activity on its platform: this applies to services of all designs, including services which are end-to-end encrypted. A service cannot design itself out of complying with the Act's illegal content safety duties.

Ofcom's powers:

Ofcom is the existing telecommunications regulator in the UK, and has now taken on additional functions as regulator of the UK's new online safety regime. It will produce a range of guidance for companies to assist compliance with the regime – the causes and impacts of harms; how services should assess and mitigate risks; identification of illegal content and Ofcom's approach to enforcement.

Where necessary and proportionate to do so, Ofcom has the power to issue a Notice to regulated providers to use accredited technology to identify and remove CSEA on any public or private part of the service. Relating to private communications, if there is no accredited technology available which is compatible with the design of a service, Ofcom can require the company to use best endeavours to develop or source such a technology. This means that where technologies are compatible with a service or platform are available, Ofcom will be able to require their use, subject to strict privacy safeguards.

Ofcom will work with companies to support compliance with the online safety regime and has a range of enforcement powers, including imposing substantial fines of up to £18m or 10% of global revenue (whichever is higher), requiring companies to make improvements and applying for business disruption measures, including blocking, via the courts.

Key summary points:

- The Online Safety Act balances the aim to keep users, particularly children, safe online, while protecting freedom of expression, and maintaining support for an innovative tech sector.
- The Act's duties and Ofcom's enforcement powers are designed to be proportionate and reasonable; building on, and formalising, safety systems and processes which in some cases services already employ to prevent illegal activity and harm on their platforms.

Mothers' Views on

KIDS ONLINE SAFETY ACT

Report of Findings

For consideration at the Hearing:

Big Tech and the Online Child Sexual Exploitation Crisis

Wednesday, January 31, 2024

Mothers Count LLC

9854 National Blvd., #526

Los Angeles, CA 90034

Table of Contents

Introduction 3

Actionable Insights at a Glance 4

Findings 6

Introduction

In October 2023 and November 2023, Count on Mothers conducted research to explore mothers' views related to the Kids Online Safety Act (KOSA). In total, 263 Mothers from 43 states completed a survey about this bill, and 7 Mothers from 7 states participated in a focus group. Actionable Insights are included in this report, followed by findings from our research. We conduct these studies because we believe that mothers' first-hand experiences and knowledge are critical sources of information in the federal policy making process, particularly on bills that affect kids or the conditions of raising a family in the United States. We share this data with legislators and the public so that they may have a better understanding of how a bill could help or hurt a family. The full report is posted on our website, located at: https://www.countonmothers.org/.

Actionable Insights at a Glance from Quantitative Study

From Mothers We Surveyed

This bill contains points of alignment with some differences among Mothers.

Overwhelmingly, Mothers were aligned across the political spectrum on the following:

- The federal government should require that social media companies provide minors with ways to protect their information, remove addictive product components, and opt out of product components that cause addiction.
- Parents should have new methods to identify harmful practices toward children and report these harms to the social media companies.
- The federal government should mandate social media companies prevent and reduce harm to minors (e.g., suicide prevention, sexual exploitation).
- Social media companies should be required to submit to a yearly independent audit that evaluates the companies' risk to minors, compliance with the law, and harm prevention efforts.
- The federal government should allow universities and public interest groups access to social media companies' databases to conduct research on harm, safety, and wellbeing to minors.
- A majority of Mothers believe that the Kids Online Safety Act will have a positive impact on the safety, health, and well-being of kids and families, including 71% of very conservative and conservative Mothers, 91% of moderate Mothers, and 89% of liberal and very liberal Mothers.

Actionable Insights at a Glance from Qualitative Study

From Mothers Who Participated in the Focus Group

- Similar themes arose as the 7 Mothers from across the political spectrum discussed their views on KOSA. Mothers are concerned that the internet, in particular social media, is having detrimental impacts on their children and families. They believe that social media companies and the federal government need to take much more responsibility for protecting their children online. They support KOSA and they would like to make sure that the legislation includes more specifics so that social media companies are held accountable for enacting the components of the legislation.

- Mothers are noticing that their children's health and mental health are being hurt by social media. Along with other parents in their communities, they are observing an increase in their children's depression, social anxiety, self-esteem problems, and bullying due to social media. This is having negative consequences on the family as a whole as well, including their ability to be effective Mothers.

- During the focus group, Mothers stated that social media companies must protect the privacy of their children, remove addictive components, allow researchers to study the effects of these technologies, and enact policies to prevent and reduce risk to minors.

- Finally, these 7 Mothers believe KOSA would create much-needed accountability for social media companies and the government, and would have a positive impact on the mental health and wellbeing of their children.

Takeaway Theme from Qualitative Study:
Social Media is Negatively Impacting
Children and Families

While they noted that social media has its benefits, Mothers discussed at length their concerns regarding its negative effects. They described how social media is causing harm to their families, in particular the wellbeing of their children. A Mother who identified as moderate commented:

"My son in particular, he watches YouTube shorts and he seems to lose track of time and just keeps watching and watching until one of us is like, please get off. I've seen that and I've seen some friends' kids getting bullied on social media, getting depressed, feeling left out and isolated. I've also seen positives, you know, people feeling more connected during the pandemic and staying in touch with people."

A conservative Mother shared similar concerns about the ways in which social media impacts children's health, and stated that these stresses are also affecting her wellbeing:

"So bad for their nervous system for their little bodies to constantly be getting that dopamine hit all the time and need it. It just dysregulates them... it completely just dysregulates their bodies and that's not even talking about electromagnetic waves and all of the radiation that they're getting from holding their phone all the time and being in front of screens and Wi Fi all the time. So, to be constantly connected is doing physiological harm, psychological harm to their entire bodies, to their entire beings. And for me, I mean, I sometimes want to go live on an island somewhere because I just feel like it's so hard to escape it."

Mothers also reported that social media is having deleterious effects on their children's relationships and experiences at school. A conservative Mother described how social media may be causing her daughter harm:

"Like since probably seven or eight years old, she's been more alienated from most of the kids her age because they started having social media and TikTok and you see them in miniskirts in second grade making videos for attention and my daughter's over here like, 'I'm not cool according to them because I don't have TikTok in second grade.' And I'm like, when did this happen?"

Later in the focus group, the same conservative Mother stated:

"*I was curious too about public schooling and how they would come into play with this because my daughter when she was probably 10 or 11, she even wrote a letter to the principal about how*

she can't even make friends because everyone's phone is in their face or their faces are in their phones, right? And I'm like, when are they going to learn to build real relationships if their phones are in their faces, even at school. Like there is a rule that says they're not supposed to use it, but how come no one can enforce it? You know what I mean?"

Findings from Quantitative and Qualitative Studies of Mothers

98% of Mothers believe the federal government should require that social media companies provide minors with ways to protect their information, remove addictive product components, and opt out of product components that cause addiction.

- 90% of very conservative and conservative Mothers were in agreement, and 3% opposed this item.
- Among moderate Mothers, 97% were in favor of this issue, 3% were unsure, and no one opposed.
- 100% of liberal and very liberal Mothers agreed with this item.

Throughout the focus group, Mothers expressed their concerns about the addictive nature of social media and how this affects the wellbeing of their children. Supporting this aspect of the bill, a moderate Mother stated:

"And I, I'm worried about, you know, what this is, these companies are just getting better and better and better at making their content more and more addictive. Like every day, you know, and with AI, it's even contributing. And so I, you know, I do worry about what is this doing to everybody's brain, especially kids' brains, and that I feel like we need to have somebody involved that's not all about profiting and making money like these companies are."

97% of Mothers believe that parents should have new methods to identify harmful practices toward children and report these harms to the social media companies.

- 94% of very conservative and conservative Mothers were in agreement, and 3% opposed this item.
- Among moderate Mothers, 98% were in favor of this issue, 3% were unsure, and no one opposed.

- 97% of liberal and very liberal Mothers agreed with this item, and 2% were in disagreement.

Mothers who participated in the focus group shared similar perspectives. A liberal Mother commented:

"If they're bullied, if a friend is bullied, because I like, we need, you know, yes, parents should have new methods to identify harmful practices towards children. The children should have those methods as well to identify."

97% of Mothers believe that the federal government should mandate social media companies prevent and reduce harm to minors (e.g., suicide prevention, sexual exploitation).

- 84% of very conservative and conservative Mothers were in agreement, and 13% were unsure, and 3% opposed this item.
- Among moderate Mothers, 99% were in favor of this issue, and 1% were opposed.
- 99% of liberal and very liberal Mothers agreed with this item, and only 1% disagreed.

During the focus group, Mothers expressed concerns about the harms of social media. Most Mothers seemed to support the aspect of the bill which grants the government responsibilities for mandating that social media companies protect their children. A Mother who identified as liberal stated:

"I've heard about a lot of negative things. I've heard about, you know, young children committing suicide due to being bullied or, you know, being treated terribly online. And that is a really serious, obviously, problem. So that obviously concerns me. And I think that this bill is really a huge, huge thing because a lot of us are, some of us have different capabilities in terms of like, how much are we on it? What do we really know about it? And I feel like this has kind of been the generation of like, some of us know, some of us don't. And we're having to kind of wade through these waters of how is this affecting my child and what controls do I need to put in place? And so I feel like, you know, having the government involved at least because there's nobody monitoring these companies now."

96% of Mothers believe that social media companies should be required to submit to a

yearly independent audit that evaluates the companies' risk to minors, compliance with the law, and harm prevention efforts.

- 87% of very conservative and conservative Mothers were in agreement, 3% were unsure, and 10% opposed this item.
- Among moderate Mothers, 98% were in favor of this issue, 2% were unsure, and no one opposed.
- 97% of liberal and very liberal Mothers agreed with this item, 2% were unsure, and only 1% disagreed.

A liberal Mother from the focus group stated that social media companies, rather than Mothers, should be primarily responsible for routinely ensuring that online content is safe for children:

"They are the ones making money. They are the ones putting our children at risk. We should not have to just be the bad guy all the time and say, my child's gonna be a social outcast because of my rules and everybody else is doing it. There should be protections that are good for all children just the way we have age limits on other controlled substances and content."

86% of Mothers believe that the federal government should allow universities and public interest groups access to social media companies' databases to conduct research on harm, safety, and wellbeing to minors.

- 68% of very conservative and conservative Mothers were in agreement, 6% were unsure, and 26% opposed this item.
- Among moderate Mothers, 86% were in favor of this issue, 14% were unsure, and no one opposed.
- 89% of liberal and very liberal Mothers agreed with this item, 9% were unsure, and only 2% disagreed.

Most mothers who participated in the focus group also thought that research should be conducted on the practices of social media companies. A moderate Mother stated:

"I think it's really important for the, um, there's a lot of research being done on what social media is doing to like the, to brains and the addictive piece of it. And I, you know, I know there are, you know, universities are, are doing a lot of research in this area and Common Sense Media is a really great resource that has a lot of information in this area."

Political Ideology

Count on Mothers is committed to representing Mothers of all political ideologies so that Congress can be best informed of constituents' experiences and views. In the overall sample of KOSA survey participants, 5% were very conservative, 7% were conservative, 35% were moderate, 38% were liberal, and 13% were very liberal.

Regarding focus group participants, 29% identified as conservative, 29% identified as moderate, and 43% identified as liberal.

State of Residence

Mothers who responded to the KOSA survey resided in 43 states. The states that had the most representation were California (41), New York (19), Connecticut (15), Virginia (15), and Illinois (12). Mothers who participated in the focus group were from the states of California, Illinois, New York, Ohio, South Carolina, and Texas.

Phyllis A. Jones (*pro hac vice*)
COVINGTON & BURLING LLP
One CityCenter
850 Tenth Street, NW
Washington, DC 20001-4956
Telephone: + 1 (202) 662-6000
Facsimile: + 1 (202) 662-6291
Email: pajones@cov.com

Attorneys for Defendants Meta Platforms, Inc., Instagram, LLC, Meta Payments, Inc., Meta Platforms Technologies, LLC, Facebook Payments, Inc., Siculus, Inc., Facebook Operations, LLC, and Mark Elliot Zuckerberg

Additional parties and counsel listed on signature pages

UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

OAKLAND DIVISION

IN RE: SOCIAL MEDIA ADOLESCENT ADDICTION/PERSONAL INJURY PRODUCTS LIABILITY LITIGATION

THIS DOCUMENT RELATES TO:

ALL ACTIONS

MDL No. 3047

Case No. 4:22-md-03047-YGR-PHK

Honorable Yvonne Gonzalez Rogers

MARK ZUCKERBERG'S MOTION TO DISMISS PURSUANT TO RULE 12(b)(6) THE PERSONAL INJURY PLAINTIFFS' CLAIMS

Hearing:
Date: TBD
Time: TBD
Place: Oakland, California
Judge: Hon. Yvonne Gonzalez Rogers

NOTICE OF MOTION AND MOTION

PLEASE TAKE NOTICE THAT, at a hearing date and time to be determined in accordance with the Court's Case Management Order No. 6 (4:22-md-3047, Dkt. No. 451), before the Honorable Yvonne Gonzalez Rogers, in Courtroom 1, Floor 4, of the United States District Court, Northern District of California, located at 1301 Clay Street in Oakland, California, Defendant Mark Zuckerberg will and hereby does move this Court, under Federal Rule of Civil Procedure 12(b)(6), for an order dismissing with prejudice Counts 8 and 9 of the Personal Injury Second Amended Master Complaint (Dkt. 494) against him in their entirety.

This Motion is based on the Memorandum of Points and Authorities submitted herewith, any Reply Memorandum or other papers submitted in connection with the Motion, the Motion to Dismiss concurrently filed by the Meta Defendants (which Mr. Zuckerberg also joins), the Personal Injury Second Amended Master Complaint filed in this action, any matter of which this Court may properly take judicial notice, and any information presented at argument.

DATED: December 22, 2023 By: */s/ Phyllis A. Jones*

Phyllis A. Jones

Attorneys for Defendants Defendants Meta Platforms, Inc., Instagram, LLC, Meta Payments, Inc., Meta Platforms Technologies, LLC, Facebook Payments, Inc., Siculus, Inc., Facebook Operations, LLC, and Mark Elliot Zuckerberg

Additional counsel listed on signature pages

TABLE OF CONTENTS

Page

I. INTRODUCTION 1

II. BACKGROUND 1

A. Mark Zuckerberg Is the Founder and CEO of Meta, Which Owns and Operates Facebook and Instagram 1

B. Plaintiffs' Claims Against Mr. Zuckerberg 2

III. LEGAL STANDARD 3

IV. ARGUMENT 4

A. Plaintiffs Must Allege Facts that Would Support Holding Mr. Zuckerberg Liable in His Personal Capacity. 4

B. Plaintiffs Fail to State a Claim for Fraud Against Mr. Zuckerberg. 6

1. Plaintiffs Fail to Plead Actual Reliance on Mr. Zuckerberg's Statements or Injury Resulting Therefrom. 7

2. Mr. Zuckerberg's Alleged Statements Consist of Non-Actionable Statements of Opinion or Are Indisputably True. 9

3. Mr. Zuckerberg's Statements to Congress Are Protected By the First Amendment. 11

4. Plaintiffs Do Not Allege Mr. Zuckerberg Had an Independent Duty to Disclose. 12

V. CONCLUSION 13

TABLE OF AUTHORITIES

CASES Page(s)

Allied Tube & Conduit Corp. v. Indian Head, Inc.,
486 U.S. 492 (1988) 11

Amado v. Procter & Gamble Co.,
2023 WL 3898984 (N.D. Cal. June 8, 2023) 11

Aprigliano v. American Honda Motor Co., Inc.,
979 F. Supp. 2d 1331, 81 U.C.C. Rep. Serv. 2d 1101 (S.D. Fla. 2013) 12

Ashcroft v. Iqbal,
556 U.S. 662 (2009) 3, 8

Azoulai v. BMW of N. Am. LLC,
2017 WL 1354781 (N.D. Cal. Apr. 13, 2017) 9

Brock v. Zuckerberg,
2021 WL 2650070 (S.D.N.Y. June 25, 2021) 5, 6

In re Brookhaven National Laboratory Trichloroethylene Cases,
511 F. Supp. 3d 374, 108 Fed. R. Serv. 3d 757 (E.D. N.Y. 2020) 12

Bruce v. Clark Equip. Co.,
2007 WL 926530 (E.D. Cal. Mar. 27, 2007) 5

Castaneda v. Amazon.com, Inc.,
--- F. Supp. 3d ---, 2023 WL 4181275 (N.D. Ill. June 26, 2023) 9

Cooke v. Allstate Mgmt. Corp.,
741 F. Supp. 1205 (D.S.C. 1990) 10

Crigler v. Salac,
438 So. 2d 1375 (Ala. 1983) 5

Cullen v. Netflix, Inc.,
2013 WL 140103 (N.D. Cal. Jan. 10, 2013) 11

E.R.R. Presidents Conf. v. Noerr Motor Freight, Inc.,
365 U.S. 127 (1961) 11

Elias v. Hewlett-Packard Co.,
903 F. Supp. 2d 843 (N.D. Cal. 2012) 9

Erickson v. Bos. Sci. Corp.,
846 F. Supp. 2d 1085 (C.D. Cal. 2011) 4

In re Fluoroquinolone Prod. Liab. Litig.,
517 F. Supp. 3d 806 (D. Minn. 2021) 12

In re Ford Motor Co. Sec. Litig.,
381 F.3d 563 (6th Cir. 2004) 9

In re Fresenius Gramuflo/Naturalyte Dialysate Prod. Liab. Litig.,
76 F. Supp. 3d 321 (D. Mass. 2015) 5

Garcia v. Chrysler Grp. LLC,
127 F. Supp. 3d 212 (S.D.N.Y. 2015) 12

Greater Hous. Transp. Co. v. Uber Techs., Inc.,
155 F. Supp. 3d 670 (S.D. Tex. 2015) 9

Haskins v. Symantec Corp.,
654 Fed. Appx. 338 (9th Cir. 2016) 7

Kearns v. Ford Motor Co.,
567 F.3d 1120 (9th Cir. 2009) 4

Kottle v. Nw. Kidney Centers,
146 F.3d 1056 (9th Cir. 1998) 11

Lloyd v. Facebook, Inc.,
2022 WL 4913347 (N.D. Cal. Oct. 3, 2022) 5, 6

In re Lyft Inc. Sec. Litig.,
484 F. Supp. 3d 758 (N.D. Cal. 2020) 9

Macpherson v. Eccleston,
190 Cal. App. 2d 24 (1961) 6

McCabe v. Daimler AG,
160 F. Supp. 3d 1337 (N.D. Ga. 2015) 12

Mewawalla v. Middleman,
601 F. Supp. 3d 574 (N.D. Cal. 2022) 4

Morris v. Princess Cruises, Inc.,
236 F.3d 1061 (9th Cir. 2001) 9

Newcal Indus., Inc. v. Ikon Off. Sol.,
513 F.3d 1038 (9th Cir. 2008) 9

O'Connor v. Uber Techs., Inc.,
2013 WL 6354534 (N.D. Cal. Dec. 5, 2013) 5

In re Optical Disk Drive Antitrust Litig.,
2011 WL 3894376 (N.D. Cal. Aug. 3, 2011) 3

Patino v. Cnty. of Monterey,
2023 WL 375349 (N.D. Cal. Jan. 24, 2023) 5

Perkins v. LinkedIn Corp.,
53 F. Supp. 3d 1190 (N.D. Cal. 2014) 8

Pirozzi v. Apple Inc.,
913 F. Supp. 2d 840 (N.D. Cal. 2012) 7

Song v. Champion Petfoods USA, Inc.,
2020 WL 7624861 (D. Minn. Dec. 22, 2020) 9

Sonora Diamond Corp. v. Superior Ct.,
83 Cal. App. 4th 523 (2000) 6

Spencer v. Hartford Fin. Servs. Grp., Inc.,
256 F.R.D. 284 (D. Conn. 2009) 7, 8

Swartz v. KPMG LLP,
476 F.3d 756 (9th Cir. 2007) 4

Tabler v. Panera LLC,
2019 WL 5579529 (N.D. Cal. Oct. 29, 2019) 7

In re Takata Airbag Prod. Liab. Litig.,
---F. Supp. 3d.---, 2023 WL 4925368 (S.D. Fla. June 20, 2023) 7

In re Takata Airbag Products Liability Litigation,
193 F. Supp. 3d 1324 (S.D. Fla. 2016) 12

United Mine Workers of Am. v. Pennington,
381 U.S. 657 (1965) 11

United States v. Bestfoods,
524 U.S. 51 (1998) 6

Vess v. Ciba-Geigy Corp. USA,
317 F.3d 1097 (9th Cir. 2003) 4

In re Welding Fume Prod. Liab. Litig.,
2007 WL 1087605 (N.D. Ohio Apr. 9, 2007) 7, 8

XYZ Two Way Radio Serv., Inc. v. Uber Techs., Inc.,
214 F. Supp. 3d 179 (E.D.N.Y. 2016) 10

In re Yahoo! Inc. Customer Data Sec. Breach Litig.,
2017 WL 3727318 (N.D. Cal. Aug. 30, 2017) 10

In re ZF-TRW Airbag Control Units Prod. Liab. Litig.,
601 F. Supp. 3d 625 (C.D. Cal. 2022) 7, 8

In re Zofran (Ondansetron) Prod. Liab. Litig.,
2017 WL 1458193 (D. Mass. Apr. 24, 2017) 3

OTHER AUTHORITIES

37 Am. Jur. 2d Fraud and Deceit § 65 9

37 Am. Jur. 2d Fraud and Deceit § 106 10

37 Am. Jur. 2d Fraud and Deceit § 238 7

Restatement (Second) of Torts § 537 cmt. a 7

I. INTRODUCTION

In Meta's concurrently filed motion to dismiss Counts 8 and 9 of the Personal Injury Second Amended Master Complaint ("PI SAC"), Meta explains why the claims for fraudulent and negligent misrepresentation and concealment fail as a matter of law. Some individual Plaintiffs in their short-form complaints have also pled these claims against Mark Zuckerberg, Meta's founder and CEO, in his personal capacity.[1] For the reasons explained in Meta's brief, as well as the additional reasons set forth below, the claims asserted against Mr. Zuckerberg fail as a matter of law.

First, it is a fundamental principle that individuals cannot be held personally liable for the acts or omissions of a corporation based merely on their status as corporate executives, directors, or shareholders. *Second*, Plaintiffs fail to allege (i) reliance on Mr. Zuckerberg's statements or omissions or (ii) injury resulting from that reliance. *Third*, none of the statements by Mr. Zuckerberg in the Complaint are actual misstatements of fact. *Fourth*, Mr. Zuckerberg cannot be civilly liable to third parties for his congressional testimony, which is protected by the First Amendment. *Fifth*, to the extent Plaintiffs allege Mr. Zuckerberg is liable for an alleged omission, they have not pled any relationship between Plaintiffs and Mr. Zuckerberg that would create a duty to disclose—a threshold ground on which such claims fail.

II. BACKGROUND

A. Mark Zuckerberg Is the Founder and CEO of Meta, Which Owns and Operates Facebook and Instagram.

Mr. Zuckerberg is the founder and Chief Executive Officer of Meta. *See* PI SAC ¶¶ 186–187. Mr. Zuckerberg founded what would become Facebook in 2003 as an undergraduate at Harvard University. *Id.* Today, Facebook is one of the world's most used social networking services, encompassing personal and professional profiles, business and creator pages, and shared-interest groups, all featuring content created by Facebook's almost three billion monthly users. *Id.* ¶¶ 187, 190. To access the Facebook service, users must register for an account. *See id.* ¶¶ 240–41. At the time of registration, users must provide their birthday and verify that they are at least 13 years old. *Id.* ¶¶ 329, 331–32, 369(c). If a

[1] Only Counts 8 and 9 are pled against Mr. Zuckerberg personally.

potential user discloses that they are not yet 13 on the sign-up page, "they are informed that they cannot create an account." *Id.* ¶ 332. Facebook will remove users that it determines are under 13. *See id.* ¶ 337.

In 2012, Meta (then Facebook) acquired Instagram. *Id.* ¶ 211. Instagram is a communication service on which billions of users worldwide create, edit, and share photos, videos, and other content. *Id.* ¶¶ 181, 210, 216, 239. Like Facebook, users must register for an account to access the service. *See id.* ¶¶ 240-41. As of 2019, potential new users must provide their birthday to create an Instagram account. *Id.* ¶ 330. Potential U.S. users that disclose they are under 13 are informed that they cannot create an account. *Id.* ¶ 332. Instagram will remove users that it determines are under the age of 13. *Id.* ¶ 337.

B. Plaintiffs' Claims Against Mr. Zuckerberg

On February 14, 2023, Plaintiffs filed their Master Complaint alleging eighteen causes of action (Dkt. 136). On December 15, 2023, Plaintiffs filled a Second Amended Master Complaint (Dkt. 494), which alleges the same eighteen causes of action. To date, 24 Plaintiffs have filed short form complaints naming Mark Zuckerberg as a Defendant in his personal capacity. Only those complaints are the subject of this motion. Those complaints assert against Mr. Zuckerberg ***only*** either Count 8 (Fraudulent Concealment and Misrepresentation) and/or Count 9 (Negligent Concealment and Misrepresentation).

The Complaint's allegations regarding Mr. Zuckerberg specifically are narrow. The vast majority of the allegations involving Meta do not mention Mr. Zuckerberg at all. *See, e.g.*, *id.* ¶¶ 181-182, 184, 188-194, 196-208, 214, 216-260, 262-270, 272-338, 340-364, 366-368, 379-387, 389-410, 412-436. Of the allegations that ***do*** mention Mr. Zuckerberg, the majority relate to alleged statements made by Mr. Zuckerberg regarding the "safety" of Meta's platforms. The PI SAC also contains several allegations regarding opinions Mr. Zuckerberg has allegedly expressed and/or aspirational statements regarding the company. Attached hereto as Appendix A is a list of each allegedly misleading statement attributed to Mr. Zuckerberg in the PI SAC.[2]

[2] In their newly Second Amended Master Complaint, the Personal Injury Plaintiffs also incorporate by reference certain allegations from the Attorneys General's Multistate Complaint. *See* Compl. ¶ 391A. The additional statements attributed to Mr. Zuckerberg in that complaint fail for the reasons explained in Meta's motion to dismiss the Attorneys General's claims, filed contemporaneously herewith.

None of the short form complaints assert factual allegations in support of their claims beyond those contained in the PI SAC. To the extent the short form complaints contain additional allegations at all, they only include the boilerplate allegation that "In Zuckerberg's testimony before Congress and in other public statements alleged in paragraphs 364 through 391 of the Master Complaint, Defendants Meta and Zuckerberg disclosed some facts but intentionally failed to disclose other facts, making their disclosures deceptive. In addition, Meta and Zuckerberg intentionally failed to disclose certain facts that were known only to them, which Plaintiff and their parents could not have discovered." *See, e.g.*, *Baker* SFC, Ex. A (4:23-cv-01578).[3]

III. LEGAL STANDARD

On Rule 12(b)(6) motions, claims are subject to dismissal for failure to plead "sufficient factual matter, accepted as true, to 'state a claim to relief that is plausible on its face.'" *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)). This well-established pleading standard applies equally in an MDL proceeding. *See In re Zofran (Ondansetron) Prod. Liab. Litig.*, 2017 WL 1458193, at *5 (D. Mass. Apr. 24, 2017) ("The creation of an MDL proceeding does not suspend the requirements of the Federal Rules of Civil Procedure, nor does it change or lower the[m]."); *In re Optical Disk Drive Antitrust Litig.*, 2011 WL 3894376, at *9 (N.D. Cal. Aug. 3, 2011) (dismissing consolidated complaint in MDL for failure to satisfy Rule 8 where plaintiffs' allegations were merely "possible" rather than "plausible.").

[3] *See also B.B.* SFC (4:23-cv-03032) (similar); *Booker* SFC (4:23-cv-01537) (similar); *C.G.* SFC (4:23-cv-01568) (similar); *C.S.* SFC (4:23-cv-01569) (similar); *Calvoni* SFC (4:22-cv-05873) (similar); *Cameron* SFC (4:23-cv-03266) (similar); *Cusato* SFC (4:23-cv-04961) (similar); *D.S.* SFC (4:23-cv-03402) (similar); *Dodd* SFC (4:23-cv-01583) (similar); *Dowdy* SFC (4:23-cv-01866) (similar); *Dyer* SFC (4:23-cv-01567) (similar); *Garceau* SFC (4:23-cv-04962) (similar); *H.D.* SFC (4:23-cv-01425) (similar); *Haas* SFC (4:23-cv-01565) (similar); *Hirka* SFC (4:23-cv-03906) (similar); *J.F.* SFC (4:23-cv-01846) (similar); *Jackson* SFC (4:23-cv-03774) (similar); *Jansky* SFC (4:23-cv-02026); *K.C.* SFC (4:23-cv-03179) (similar); *Keizer* SFC (4:23-cv-02972) (similar); *Koizol* SFC (4:23-cv-02244) (similar); *M.C.* SFC (4:23-cv-03398); *M.M.* SFC (4:23-cv-01615) (similar); *M.W.* SFC (4:23-cv-03824) (similar); *N.K.* SFC (4:23-cv-01584) (similar); *S.S.* SFC (4:23-cv-02024) (similar). Several of these short form complaints identify the district in which Plaintiff "would have filed in the absence of direct filing." Mr. Zuckerberg does not hereby concede that those courts would have personal jurisdiction over him, and reserves all rights regarding any future motion to transfer.

Because Counts 8 and 9 "sound in fraud," they are subject to the heightened pleading standards of Rule 9(b). *Vess v. Ciba-Geigy Corp. USA*, 317 F.3d 1097, 1103 (9th Cir. 2003). To plead these fraud-based claims with the particularity required by Rule 9(b), Plaintiffs must plead the "time, place, and specific content of the false representations as well as the identities of the parties to the misrepresentations." *Swartz v. KPMG LLP*, 476 F.3d 756, 764 (9th Cir. 2007). Rule 9(b) also applies to omission-based claims. *See Kearns v. Ford Motor Co.*, 567 F.3d 1120, 1127 (9th Cir. 2009) ("[N]ondisclosure is a claim for misrepresentation in a cause of action for fraud, [so] it (as any other fraud claim) must be pleaded with particularity under Rule 9(b)."); *Erickson v. Bos. Sci. Corp.*, 846 F. Supp. 2d 1085, 1093 (C.D. Cal. 2011) (omission claims "must describe the content of the omission and where the omitted information should or could have been revealed").

IV. ARGUMENT

Plaintiffs' claims against Mr. Zuckerberg fail because they have not alleged any facts that would support a fraudulent or negligent misrepresentation/omission claim against him. As a matter of law, Mr. Zuckerberg cannot be held personally liable for fraudulent or negligent misrepresentation or omissions claims based on alleged misstatements or omissions that are attributed to others at the company or the company in general. Even for representations attributed specifically to Mr. Zuckerberg, the claims fail on the merits because Plaintiffs do not allege actual reliance, injury, or actionable misstatements of fact. And Mr. Zuckerberg's congressional testimony, which is cited liberally in the PI SAC, is protected by the First Amendment. Finally, to the extent Plaintiffs' claims are based on alleged omissions, Plaintiffs have not pled any facts that would create a duty to disclose. For the reasons set forth herein, as well as for the additional reasons set forth in Meta's concurrently filed motion to dismiss, which is incorporated by reference herein, the claims against Mr. Zuckerberg should be dismissed in their entirety.

A. Plaintiffs Must Allege Facts that Would Support Holding Mr. Zuckerberg Liable in His Personal Capacity.

As a preliminary matter, Plaintiffs cannot hold Mr. Zuckerberg liable for any conduct Meta allegedly engaged in merely because he is the CEO of Meta. It is black letter law that, to hold a corporate officer personally liable, "a plaintiff must first show that the director ***specifically authorized, directed or participated*** in the allegedly tortious conduct." *Mewawalla v. Middleman*, 601 F. Supp. 3d 574, 597 n.2

(N.D. Cal. 2022) (emphasis added) (citing *Frances T. v. Vill. Green Owners Assn.*, 723 P.2d 573 (Cal. 1986)); *see also, e.g.*, *Crigler v. Salac*, 438 So. 2d 1375, 1380 (Ala. 1983) ("In order to hold an officer of a corporation liable for the negligent or wrongful acts of the corporation . . . he must be a participant in the wrongful act.").

Mere status as an executive does not confer liability on a corporate officer for alleged conduct of the corporation. And courts routinely grant motions to dismiss corporate officers from suits where plaintiffs failed to level specific allegations that would support the corporate officer's direct participation in the allegedly tortious conduct. *See, e.g.*, *O'Connor v. Uber Techs., Inc.*, 2013 WL 6354534, at *18 (N.D. Cal. Dec. 5, 2013) (dismissing claims against executive officers of Uber because plaintiffs failed to allege enough specific allegations demonstrating that the executives "personally directed or participated in the tortious conduct."); *Patino v. Cnty. of Monterey*, 2023 WL 375349, at *7 (N.D. Cal. Jan. 24, 2023) (dismissing claims against company executives because plaintiff did not specifically identify what conduct was attributable to each of the executives as opposed to the corporation itself); *In re Fresenius Gramuflo/Naturalyte Dialysate Prod. Liab. Litig.*, 76 F. Supp. 3d 321, 336 (D. Mass. 2015) (dismissing claims against a corporate officer because plaintiffs failed to make specific allegations that would support holding the officer individually liable).

Indeed, multiple courts have dismissed Mr. Zuckerberg specifically from suits naming him as a defendant in his personal capacity where plaintiffs failed to make allegations that would support holding him personally liable. *See Lloyd v. Facebook, Inc.*, 2022 WL 4913347, at *5 (N.D. Cal. Oct. 3, 2022) (dismissing various claims, including fraud, against Mr. Zuckerberg because the complaint lacked specific factual allegations to state a plausible claim for his personal liability); *Brock v. Zuckerberg*, 2021 WL 2650070, at *4 (S.D.N.Y. June 25, 2021); (dismissing individual claims against Mr. Zuckerberg and another Meta executive because the complaint lacked allegations that would support holding them personally liable for the alleged conduct of the company).

Nor can Plaintiffs seek to hold Mr. Zuckerberg liable based solely on his status as a shareholder in the company. "It is a fundamental rule of corporate formation that a shareholder, be it another corporation or an individual, ***is not liable for the actions of the corporation***." *Bruce v. Clark Equip. Co.*, 2007 WL 926530, at *3 (E.D. Cal. Mar. 27, 2007) (emphasis added) (citing *Mesler v. Bragg Mgmt. Co.*, 39 Cal.3d

290, 300 (1985)); *see also United States v. Bestfoods*, 524 U.S. 51, 61 (1998) ("It is a general principle of corporate law deeply 'ingrained in our economic and legal systems' that a parent corporation (so-called because of control through ownership of another corporation's stock) is not liable for the acts of its subsidiaries."). Courts will only "pierce the corporate veil" to hold a shareholder liable for the acts of a corporation in exceptional cases, requiring proof that the corporation is the "alter ego" or "instrumentality" of the individual (or parent company) in question. *See, e.g.*, *Sonora Diamond Corp. v. Superior Ct.*, 83 Cal. App. 4th 523, 539 (2000) ("Alter ego is an extreme remedy, sparingly used.").[4] Here, Plaintiffs do not allege that Meta is the "alter ego" of Mr. Zuckerberg—nor could they. *See Lloyd*, 2022 WL 4913347 at *5 (granting motion to dismiss where plaintiffs alleged Mr. Zuckerberg was liable under an "alter ego" theory).

In short, to state a claim against Mr. Zuckerberg under Counts 8 and/or 9, they cannot rely on allegations regarding Meta's corporate conduct generally.[5] Instead, Plaintiffs must make allegations ***specific to Mr. Zuckerberg*** that state a claim under either a fraudulent or negligent misrepresentation/omission theory. For the reasons discussed below, Plaintiffs fail to do so here.

B. Plaintiffs Fail to State a Claim for Fraud Against Mr. Zuckerberg.

For four separate reasons, Plaintiffs fail adequately to plead negligent or fraudulent misrepresentation or omission claims based on the statements or omissions attributed to Mr. Zuckerberg. First, Plaintiffs fail to plead actual reliance on the alleged statements or omissions. Second, the alleged statements are generalized statements of opinion, not actionable statements of fact. Third, many of the statements were made to Congress and are protected by the First Amendment. And, finally, Mr. Zuckerberg owes no duty to Plaintiffs, as required for claims based on alleged omissions.

[4] *See also Macpherson v. Eccleston*, 190 Cal. App. 2d 24, 27 (1961) ("The corporate entity may be disregarded only when there is 'such unity of interest and ownership that the separate personalities of the corporation and the individual no longer exist'").

[5] Plaintiffs also fail to state a claim under Counts 8 and 9 against Meta for the reasons stated in Meta's concurrently filed motion to dismiss these same counts.

1. Plaintiffs Fail to Plead Actual Reliance on Mr. Zuckerberg's Statements or Injury Resulting Therefrom.

Plaintiffs fail to plead actual reliance on Mr. Zuckerberg's alleged misrepresentations and omissions, as required.

"[T]he fundamental elements of fraud are substantially similar from state to state." *In re Takata Airbag Prod. Liab. Litig.*, ---F. Supp. 3d.---, 2023 WL 4925368, at *12 (S.D. Fla. June 20, 2023). "Virtually every state requires that . . . the plaintiff relied on the statement, and the plaintiff was injured as a result." *Spencer v. Hartford Fin. Servs. Grp., Inc.*, 256 F.R.D. 284, 301 (D. Conn. 2009). This is true irrespective of whether the claim is based on an alleged statement or omission, and irrespective of whether it is pled under a theory of fraud or negligent misrepresentation.[6] Additionally, reliance must be pled in accordance with the requirements of Rule 9(b). *E.g.*, *Haskins v. Symantec Corp.*, 654 Fed. Appx. 338, 339 (9th Cir. 2016) ("Because Haskins's complaint did not allege that she read and relied on a specific misrepresentation by [defendant], she failed to plead her fraud claims with particularity as required by Rule 9(b).").[7]

No Plaintiff has alleged actual reliance on any of the specific representations allegedly made by Mr. Zuckerberg—let alone with the particularity required by Rule 9(b). No Plaintiff has alleged that he or she even saw any of the supposedly false statements, or the statements in which Plaintiffs allege a

---

[6] *See* Restatement (Second) of Torts § 537 cmt. a ("The recipient of a fraudulent misrepresentation can recover from the maker for his pecuniary loss only if he in fact relies upon the misrepresentation in acting or in refraining from action, and his reliance is a substantial factor in bringing about the loss."); 37 Am. Jur. 2d Fraud and Deceit § 238 ("[W]here redress is sought for fraudulent concealment, it must appear that the party seeking relief relied upon the one with whom he or she was doing business to disclose the true facts and circumstances relating to the transaction and that the suppression of such facts was an inducement which moved him or her to enter into the agreement."); *see also In re Welding Fume Prod. Liab. Litig.*, 2007 WL 1087605, at *7 (N.D. Ohio Apr. 9, 2007) ("reasonable reliance" an element of claim for negligent misrepresentation).

[7] *See Tabler v. Panera LLC*, 2019 WL 5579529, at *12 (N.D. Cal. Oct. 29, 2019) ("[A] plaintiff does not satisfy Rule 9(b) when the plaintiff generally identifies allegedly misleading statements but fails to specify which statements the plaintiff actually saw and relied upon."); *Pirozzi v. Apple Inc.*, 913 F. Supp. 2d 840, 850 (N.D. Cal. 2012) (requiring under Rule 9(b) plaintiff to "specify when she was exposed to the statements or which ones she found material to her decisions to purchase"); *see also In re ZF-TRW Airbag Control Units Prod. Liab. Litig.*, 601 F. Supp. 3d 625, 767 (C.D. Cal. 2022) (requiring plaintiffs "to plead some facts 'to establish that they would have been aware of the [omitted fact], if it were disclosed.'" (citation omitted))

disclosure should have been made. *See Perkins v. LinkedIn Corp.*, 53 F. Supp. 3d 1190, 1220 (N.D. Cal. 2014) ("To make the reliance showing, . . . plaintiffs in misrepresentation cases must allege that they actually read the challenged representations."); *In re ZF-TRW Airbag Control Units Prod. Liab. Litig.*, 601 F. Supp. 3d 625, 767 (C.D. Cal. 2022) (omission claim failed to allege reliance because it did not "allege that Plaintiffs were exposed to the materials that should have included a disclosure of the Alleged Defect, or that they would have seen it had it been disclosed").

None of the Plaintiffs pleads any facts supporting actual reliance. They assert only legal conclusions—*e.g.*, "Had the omitted information been disclosed, the injuries that Plaintiff suffered would have been avoidable and avoided," *Baker* SFC, Ex. A (4:23-cv-01578).[8] On a motion to dismiss, the court should not credit those purely conclusory legal allegations. *Iqbal*, 556 U.S. at 678.[9] Accordingly, the absence of any well-pled allegations regarding actual reliance requires the dismissal of the claims against Mr. Zuckerberg.

Relatedly, Plaintiffs also fail to plausibly plead "injury resulting to [them] because of such reliance." *In re Welding Fume Prod. Liab. Litig.*, 2007 WL 1087605, at *7; *accord Spencer v. Hartford Fin. Servs. Grp., Inc.*, 256 F.R.D. 284, 301 (D. Conn. 2009) ("Virtually every state requires that . . . the plaintiff relied on the statement, and the plaintiff was injured as a result" of reliance on the alleged misrepresentation.). Even assuming that Plaintiffs adequately pled injury relating to their use of social media, that would not be enough. Rather, to adequately allege a claim sounding in fraud, Plaintiffs must

---

[8] *See also B.B.* SFC (4:23-cv-03032) (similar); *Booker* SFC (4:23-cv-01537) (similar); *C.G.* SFC (4:23-cv-01568) (similar); *C.S.* SFC (4:23-cv-01569) (similar); *Calvoni* SFC (4:22-cv-05873) (similar); *Cameron* SFC (4:23-cv-03266) (similar); *Cusato* SFC (4:23-cv-04961) (similar); *D.S.* SFC (4:23-cv-03402) (similar); *Dodd* SFC (4:23-cv-01583) (similar); *Dowdy* SFC (4:23-cv-01866) (similar); *Dyer* SFC (4:23-cv-01567) (similar); *Garceau* SFC (4:23-cv-04962) (similar); *H.D.* SFC (4:23-cv-01425) (similar); *Haas* SFC (4:23-cv-01565) (similar); *Hirka* SFC (4:23-cv-03906) (similar); *J.F.* SFC (4:23-cv-01846) (similar); *Jackson* SFC (4:23-cv-03774) (similar); *Jansky* SFC (4:23-cv-02026); *K.C.* SFC (4:23-cv-03179) (similar); *Keizer* SFC (4:23-cv-02972) (similar); *Koizol* SFC (4:23-cv-02244) (similar); *M.C.* SFC (4:23-cv-03398); *M.M.* SFC (4:23-cv-01615) (similar); *M.W.* SFC (4:23-cv-03824) (similar); *N.K.* SFC (4:23-cv-01584) (similar); *S.S.* SFC (4:23-cv-02024) (similar).

[9] To be sure, some Plaintiffs do plead a bit more specifically how they would have acted differently. *E.g.*, *Booker* SFC ("Had the omitted information been disclosed, Plaintiff Richard Neal Booker reasonably would have prohibited his minor child S.B. from ever downloading and using Instagram."). But without establishing how they would have actually been aware of the disclosure in the first place, the claim of reliance is not plausible and does not satisfy Rule 9(b).

allege a connection between the statements that they challenge and their alleged harms. Because Plaintiffs do not even attempt to allege facts plausibly suggesting that they were injured as a result of their reliance on Mr. Zuckerberg's statements, their claims fail as a matter of law.

2. Mr. Zuckerberg's Alleged Statements Consist of Non-Actionable Statements of Opinion or Are Indisputably True.

Plaintiffs' claims against Mr. Zuckerberg also fail because they do not adequately plead actionable misstatements of fact.

It is well established that misrepresentation claims must be based on misstatements of ***fact***, capable of being proven false. *See* 37 Am. Jur. 2d Fraud and Deceit § 65 ("The principle is fundamental that fraud cannot be predicated upon the mere expression of an opinion."). "Thus, a statement that is quantifiable, that makes a claim as to the 'specific or absolute characteristics of a product [or service],' may be an actionable statement of fact while a general, subjective claim about a product is non-actionable puffery." *Newcal Indus., Inc. v. Ikon Off. Sol.*, 513 F.3d 1038, 1053 (9th Cir. 2008). *See also, e.g.*, *Elias v. Hewlett-Packard Co.*, 903 F. Supp. 2d 843, 854–55 (N.D. Cal. 2012) ("Generalized, vague, and unspecified assertions constitute 'mere puffery' upon which a reasonable consumer could not rely, and hence are not actionable."); *Song v. Champion Petfoods USA, Inc.*, 2020 WL 7624861, at *10 (D. Minn. Dec. 22, 2020), *aff'd*, 27 F.4th 1339 (8th Cir. 2022) (statements are not actionable when they are "too vague to be proved or disproved"); *Castaneda v. Amazon.com, Inc.*, --- F. Supp. 3d ---, 2023 WL 4181275, at *7 (N.D. Ill. June 26, 2023) (statements that are "not objectively verifiable" are not actionable).

In particular, courts routinely dismiss misrepresentation claims based on statements regarding "safety" or the prioritization of safety. *See, e.g.*, *Morris v. Princess Cruises, Inc.*, 236 F.3d 1061, 1068 (9th Cir. 2001) (representation that consumers "would be safely and adequately served" failed to state a claim because the statement "is devoid of any meaningful specificity"); *In re Lyft Inc. Sec. Litig.*, 484 F. Supp. 3d 758, 770 (N.D. Cal. 2020) ("generalized assertions about Lyft's commitment to safety, its safety measures, and the role safety plays in the rideshare market" were non-actionable puffery); *Azoulai v. BMW of N. Am. LLC*, 2017 WL 1354781, at *8 (N.D. Cal. Apr. 13, 2017) ("[T]here is nothing 'specific and measurable' about the word 'safely.'"); *In re Ford Motor Co. Sec. Litig.*, 381 F.3d 563, 570 (6th Cir. 2004) (statements about "quality [and] safety" were nonactionable opinions); *Greater Hous. Transp. Co. v. Uber*

Techs., Inc., 155 F. Supp. 3d 670, 683 (S.D. Tex. 2015) (statement that Uber is the "safest ride on the road" was "non-actionable puffery"); *XYZ Two Way Radio Serv., Inc. v. Uber Techs., Inc.*, 214 F. Supp. 3d 179, 183–84 (E.D.N.Y. 2016) (similar, for statement about "mak[ing] Uber the safest experience"); *Cooke v. Allstate Mgmt. Corp.*, 741 F. Supp. 1205, 1216 (D.S.C. 1990) (a "judgment and opinion about safety" "is certainly not the kind of statement [the] law would support as fraudulent").

In re Yahoo! Inc. Customer Data Sec. Breach Litig., 2017 WL 3727318 (N.D. Cal. Aug. 30, 2017), is particularly instructive. The plaintiffs in that multi-district litigation challenged Yahoo's statement that "protecting our systems and our users' information is paramount" as misleading because Yahoo allegedly knew and did not disclose the risks of a breach of users' personal data, and then did not disclose actual data breaches after they occurred. *Id.* at *2–4, 26. The court dismissed the consumer protection claims with respect to this statement because it "is a vague and 'all-but-meaningless superlative[]' regarding how Defendants' prioritize the safety of their systems and their users' information." *Id.* at *26 (citation omitted).

Here, the majority of Mr. Zuckerberg's alleged misrepresentations constitute precisely such generalized, unmeasurable statements regarding Meta's prioritization of "safety." *See, e.g.*, Compl. ¶ 369a ("we're really focused on, on safety, especially children's safety"); *see also* Appx. A. And, with one exception, all involve vague, non-specific matters of opinion. *See, e.g.*, PI SAC ¶ 370a ("We will do our part to make this [better world] happen, not only because we love you, but also because we have a moral responsibility to all children in the next generation."); *see also* Appx. A. Such statements are simply not the sort of concrete, measurable statements of fact that are actionable in fraud.

The ***only*** concrete claim by Mr. Zuckerberg set out in the Master Complaint is the assertion that Meta "do[es] not allow people under the age of 13 to sign up" for its services. *Id.* ¶ 369c; *see also id.* ¶ 388 ("There is clearly a large number of people under the age of 13 who would want to use a service like Instagram. We currently do not allow them to do that."). But that statement is ***undisputedly true***, as Plaintiffs acknowledge in their Complaint. *See id.* ¶ 332 ("If the user reports a birthday indicating they are less than 13 years old, they are informed that they cannot create an account."). "It is fundamental in the law of fraud that a representation must be false to warrant a basis for relief." *See* 37 Am. Jur. 2d Fraud and Deceit § 106.

To be sure, Plaintiffs allege that Meta has generalized knowledge that some users lie about their age and violate Meta's policies by creating accounts before their thirteenth birthday. *See, e.g.*, PI SAC ¶ 389. But that allegation does not in any way call into question or render false the actual statements attributed to Mr. Zuckerberg. *See Amado v. Procter & Gamble Co.*, 2023 WL 3898984, at *9 (N.D. Cal. June 8, 2023) (representation is not false if there is a "mismatch between the representations at issue and the evidence that allegedly debunks them."); *Cullen v. Netflix, Inc.*, 2013 WL 140103, at *7 (N.D. Cal. Jan. 10, 2013), *aff'd*, 600 F. App'x 508 (9th Cir. 2015) ("Plaintiff has not alleged facts showing that Defendant's assertions and calculations turned out to be false" because "Plaintiff's presentation of his own calculation . . . does not contradict Defendant's claim"). Precisely such a mismatch exists here: it can be true both that Meta prohibits people under the age of 13 from using its services ***and*** that it has general awareness that some people will try to circumvent that prohibition.

3. Mr. Zuckerberg's Statements to Congress Are Protected By the First Amendment.

Several of the statements made by Mr. Zuckerberg in the Master Complaint were made in testimony to Congress. *See* PI SAC ¶¶ 369b, 369c, 370b, 370c, 370p. The First Amendment bars claims based on efforts to petition the government, including Congressional testimony. *See, e.g.*, *United Mine Workers of Am. v. Pennington*, 381 U.S. 657, 670 (1965) (efforts to influence public officials are not illegal, "regardless of intent or purpose"). The *Noerr-Pennington* doctrine applies even to allegedly false statements. *See, e.g.*, *Allied Tube & Conduit Corp. v. Indian Head, Inc.*, 486 U.S. 492, 499-500 (1988) ("[a] publicity campaign directed at the general public, seeking legislation or executive action, enjoys antitrust immunity even when the campaign employs unethical and deceptive methods"); *E.R.R. Presidents Conf. v. Noerr Motor Freight, Inc.*, 365 U.S. 127, 145 (1961) (defendant could not be liable for "attempt[ing] to bring about the passage of laws" even if it "deliberately deceived the public and public officials"); *Kottle v. Nw. Kidney Centers*, 146 F.3d 1056, 1060–62 (9th Cir. 1998) ("the sham exception is extraordinarily narrow" in the context of legislative proceedings). For this reason, Plaintiffs may not assert claims based on Mr. Zuckerberg's alleged misstatements made to Congress.

4. Plaintiffs Do Not Allege Mr. Zuckerberg Had an Independent Duty to Disclose.

Finally, Plaintiffs' claims fail to the extent they are based on any alleged "concealment" or "omission" because they do not allege any relationship with Mr. Zuckerberg that would give rise to an independent duty to disclose.

To hold an individual liable for alleged omissions, courts across jurisdictions have held that nondisclosure or concealment becomes fraudulent only when it violates a duty to disclose. *See In re Takata Airbag Products Liability Litigation*, 193 F. Supp. 3d 1324 (S.D. Fla. 2016) (applying Alabama law); *Aprigliano v. American Honda Motor Co., Inc.*, 979 F. Supp. 2d 1331, 81 U.C.C. Rep. Serv. 2d 1101 (S.D. Fla. 2013) (applying Florida law); *In re Brookhaven National Laboratory Trichloroethylene Cases*, 511 F. Supp. 3d 374, 108 Fed. R. Serv. 3d 757 (E.D. N.Y. 2020), certification denied, 514 F. Supp. 3d 546 (E.D. N.Y. 2021) (applying New York law); *Garcia v. Chrysler Grp. LLC*, 127 F. Supp. 3d 212, 237 (S.D.N.Y. 2015) ("Plaintiffs have failed to allege a duty to disclose and, by extension, a plausible fraudulent concealment claim under the laws of Alabama, Florida, Georgia, New Jersey, South Dakota, and Texas."); *McCabe v. Daimler AG*, 160 F. Supp. 3d 1337, 1350 (N.D. Ga. 2015) ("However, in a fraudulent concealment action, there must first exist a duty to communicate the omitted or concealed material fact to the defrauded party.").

Typically, a duty to disclose does not exist absent a contractual or other special relationship between the parties. *See, e.g.*, *McCabe*, 160 F. Supp. 3d at 1350 (finding no duty to disclose where defendant had "no apparent relationship with Plaintiffs"); *In re Fluoroquinolone Prod. Liab. Litig.*, 517 F. Supp. 3d 806, 820 (D. Minn. 2021) ("The 'touchstone of [the Court's] duty analysis is to ask whether a plaintiff and a defendant stood in such a relationship to one another that [Illinois] law imposed upon the defendant an obligation of reasonable conduct for the benefit of the plaintiff.'").

Here, Plaintiffs have not alleged any relationship with Mr. Zuckerberg that would create such a duty. Indeed, Plaintiffs allegations are premised only on Mr. Zuckerberg's status as CEO of Meta. *See supra* at Part IV.A. The relationship between Mr. Zuckerberg and Plaintiffs is accordingly no different than the relationship between ***any*** corporate executive and plaintiffs in a tort case in which they seek to

hold a company liable. Meta is not aware of any case holding that corporate executives have a sweeping "duty" to the public in these circumstances, and this Court should similarly reject any such theory.[10]

V. CONCLUSION

Defendant Mark Zuckerberg respectfully requests that Plaintiffs' claims against him under Counts 8 and 9 of the Amended Master Complaint be dismissed with prejudice.

Dated: December 22, 2023

Respectfully submitted,

COVINGTON & BURLING LLP

/s/ Phyllis A. Jones

Paul W. Schmidt, *pro hac vice*
pschmidt@cov.com
Phyllis A. Jones, *pro hac vice*
pajones@cov.com
Christian J. Pistilli (*pro hac vice* pending)
COVINGTON & BURLING LLP
One CityCenter
850 Tenth Street, NW
Washington, DC 20001-4956
Telephone: + 1 (202) 662-6000
Facsimile: + 1 (202) 662-6291

Emily Johnson Henn (State Bar No. 269482)
ehenn@cov.com
COVINGTON & BURLING LLP
3000 El Camino Real
5 Palo Alto Square, 10th Floor
Palo Alto, CA 94306
Telephone: + 1 (650) 632-4700
Facsimile: +1 (650) 632-4800

Attorneys for Defendants Meta Platforms, Inc., Instagram, LLC, Meta Payments, Inc., Meta Platforms Technologies, LLC, Facebook Payments, Inc., Siculus, Inc., Facebook Operations, LLC, and Mark Elliot Zuckerberg

[10] To the extent Plaintiffs claim a duty to disclose arises from any affirmative statements made by Mr. Zuckerberg, such a claim would fail because Mr. Zuckerberg's statements were *not* actionable for the reasons explained herein.

Appendix A

- "[Facebook's] controls are not just to make people feel ***safe***; it's actually what people want in the product." PI SAC ¶ 173.
- "[W]e agree, stalking isn't ***cool***; but being able to know what's going on in your friends' lives is." *Id.* ¶ 195.
- "So, we're really focused on, on ***safety***, especially children's ***safety***. So we're having folks under the age of 18 . . . we just take a lot of extra precautions for it, to make sure that it's just a ***safe environment*** for them [], to use this service that you know, the default for, for people sharing things isn't that they're sharing with everyone but that they're sharing with a smaller community . . . I think that's a lot of it. We really ***try to build a safe environment*** . . . that's gonna be the key long term." *Id.* ¶ 369a.
- "Right, and they, they feel like Facebook is this ***really secure*** place and that it's ***a hundred percent safe***, and um, we're always thinking about little and big things like that that we can do to ***keep it safe*** for, for the people who use our service." *Id.* ¶ 369b.
- "I mean, we do not allow people under the age of 13 to sign up and I think if we ever were, we would need to try to figure out a lot of ways to ***make sure that they were safe***, right, because that's just ***extremely important*** and that's just not the top of the list in terms of things for us to figure out right now." *Id.* ¶ 369c.
- "We will do our part to make this [***better world***] happen, not only because we love you, but also because we have a ***moral responsibility*** to all children in the next generation." *Id.* ¶ 370a.
- "Congressman, we have a number of measures in place to protect minors specifically. We make it so that adults can't contact minors who they - they aren't already friends with. We make it so that certain content that may be ***inappropriate*** for minors, we don't show." *Id.* ¶ 370b.
- "No . . . that's not how we talk about this or how we set up our product teams. We want our products to be ***valuable*** to people, and if they're ***valuable***, then people choose to use them." *Id.* ¶ 370c.
- "There are really two core principles at play here. There's giving people a voice, so that people can express their opinions. Then there's ***keeping the community safe***, which I think is really important." *Id.* ¶ 370d.
- "[W]e will continue to invest heavily in security and privacy because we have a ***responsibility to keep people safe***. But as I've said on past calls, we're investing so much in security that it will significantly impact our profitability." *Id.* ¶ 370e.
- "One of the most important responsibilities we have as a company is to ***keep people safe*** and stop anyone from abusing our service." *Id.* ¶ 370f.

- "What I've learned so far is that when you build services that are used by billions of people across countries and cultures, you will see all of the good humanity is capable of, and people will try to abuse those services in ever way possible. It is our responsibility to ***amplify the good and mitigate the bad***." *Id.* ¶ 370g.

- "[W]e have a ***responsibility to keep people safe*** on our services – whether from terrorism, bullying, or other threats." *Id.* ¶ 370h.

- "We ended 2018 with more than 30,000 people working on ***safety*** and security – up from 10,000 people a couple of years ago." *Id.* ¶ 370i.

- "[O]n all the content and ***safety*** security issues, there's more to do here but ***I'm proud*** of the work that we have done to get in front of a lot more of these issues." *Id.* ¶ 370j.

- "[W]e have a responsibility to ***keep people safe*** on our services." *Id.* ¶ 370k.

- "You should expect we'll do everything we can to ***keep you safe*** on our services, within the bounds of an encrypted service." *Id.* ¶ 370l.

- "***I don't believe*** [Meta's platforms harm children]. This is something that we study and we **care a lot about**; designing products that improve peoples' ***well-being is very important to us***. And what our products do is help people stay connected to people they care about, which I think is one of the most fundamental and important human things that we do, whether that's for teens or for people who are older than that." *Id.* ¶ 370p.

- "If the user reports a birthday indicating they are less than 13 years old, they are informed that they cannot create an account." *Id.* ¶ 332.

Previn Warren
MOTLEY RICE LLC
401 9th Street NW Suite 630
Washington, DC 20004
Telephone: 202-386-9610
pwarren@motleyrice.com

Attorney for Plaintiffs Baker, B.B., Booker, C.G., C.S., Calvoni, Cameron, Cusato, D.S., Dodd, Dowdy, Garceau, Haas, Hirka, J.F., Jackson, Jansky, K.C., Keizer, Koizol, M.C., M.M., N.K., Robertson, and S.S.

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

IN RE: SOCIAL MEDIA ADOLESCENT ADDICTION/PERSONAL INJURY PRODUCTS LIABILITY LITIGATION

This Document Relates to:

Baker (4:23-cv-01578);
B.B. (4:23-cv-03032);
Booker (4:23-cv-01537);
C.G. (4:23-cv-01568);
C.S. (4:23-cv-01569);
Calvoni (4:22-cv-05873);
Cameron (4:23-cv-03266);
Cusato (4:23-cv-04961);
D.S. (4:23-cv-03402);
Dodd (4:23-cv-01583);
Dowdy (4:23-cv-01866);
Garceau (4:23-cv-04962);
Haas (4:23-cv-01565);
Hirka (4:23-cv-03906);
J.F. (4:23-cv-01846);
Jackson (4:23-cv-03774);
Jansky (4:23-cv-02026);
K.C. (4:23-cv-03179);
Keizer (4:23-cv-02972);
Koizol (4:23-cv-02244);
M.C. (4:23-cv-03398);
M.M. (4:23-cv-01615);
N.K. (4:23-cv-01584);
Robertson (4:24-cv-00127);
S.S. (4:23-cv-02024).

Case No. 4:22-md-03047-YGR

MDL No. 3047

PLAINTIFFS' OPPOSITION TO MARK ZUCKERBERG'S MOTION TO DISMISS

Judge Yvonne Gonzalez Rogers

TABLE OF CONTENTS

TABLE OF AUTHORITIES ii

I. INTRODUCTION 1

II. BACKGROUND 1

A. Mark Zuckerberg's power as a decision-maker and spokesperson for Meta is unparalleled. 1

B. Mark Zuckerberg concealed evidence that Meta's platforms are not safe for youth. 2

III. LEGAL STANDARDS 4

IV. ARGUMENT 5

A. Mark Zuckerberg is personally liable for his omissions and concealment. 6

B. Plaintiffs plausibly allege Mark Zuckerberg's fraudulent omissions and concealment. 7

1. Mr. Zuckerberg had a duty to disclose the dangers Meta's products present to minors 7

2. Plaintiffs adequately plead that Mr. Zuckerberg made incomplete statements of fact 11

3. Plaintiffs relied on and were harmed by Mr. Zuckerberg's misleading statements. 12

C. Mr. Zuckerberg does not have a First Amendment right to lie to Congress. . 17

V. CONCLUSION 19

TABLE OF AUTHORITIES

Cases

ABT Bldg. Prods. Corp. v. Nat'l Union Fire Ins. Co. of Pittsburgh,
472 F.3d 99 (4th Cir. 2006) 17

Acoustic Systems, Inc. v. Wenger Corp.,
207 F.3d 287(5th Cir. 2000) 17

Allied Tube & Conduit Corp. v. Indian Head, Inc.,
486 U.S. 492 (1988) 18

Ashcroft v. Iqbal,
556 U.S. 662 (2009) 4

Azoulai v. BMW of North America LLC,
2017 WL 1354781 (N.D. Cal. Apr. 13, 2017) 12

BAC Home Loans Serv. v. Farina,
2010 Conn. Super. LEXIS 4929 (Conn. Super. Ct. June 2, 2010) 8

Bain v. Jackson,
783 F. Supp. 2d 13 (D.D.C. 2010) 8

Bank of Montreal v. Signet Bank,
193 F.3d 818 (4th Cir. 1999) 8, 10

Batis v. Dun & Bradstreet Holdings, Inc.,
2023 WL 1870057 (N.D.Cal., 2023) 17

Bays v. Hunter Sav. Ass'n,
539 F. Supp. 1020 (S.D. Ohio 1982) 10

Belville v. Ford Motor Co.,
60 F. Supp. 3d 690 (S.D.W. Va. 2014) 4

Berger v. Sec. Pac. Info. Sys., Inc.,
795 P.2d 1380 (Colo. App. 1990) 9

BP Am. Prod. Co. v. Patterson,
263 P.3d 103 (Colo. 2011) 13

Brock v. Zuckerberg,
2021 WL 2650070 (S.D.N.Y. June 25, 2021) 7

Burman v. Richmond Homes Ltd.,
821 P.2d 913 (Colo. App. 1991) 10

Coastal Abstract Serv., Inc. v. First Am. Title Ins. Co.,
173 F.3d 725 (9th Cir. 1999) 6

Coldwell Banker Whiteside Assocs. v. Ryan Equity Partners, Ltd.,
181 S.W.3d 879 (Tex. App. 2006) 8

Cope v. Metro Life Ins. Co.,
82 Ohio St.3d 426, 436 (Ohio 1998) 13

Daniel v. Ford Motor Co.,
806 F.3d 1217 (9th Cir. 2015) 13, 17

Dean v. Beckley,
2010 WL 3928650, at (D. Md. Oct. 1, 2010) 8

DiMichele v. Perrella,
120 A.3d 551 (Conn. App. Ct. 2015) 10

Donsco, Inc. v. Casper Corp.,
587 F.2d 602 (3d Cir. 1978) 6

E.R.R. Presidents Conf. v. Noerr Motor Freight, Inc.,
365 U.S. 127 (1961) 18

Edmondson & Gallagher v. Alban Towers Tenants Ass'n,
48 F.3d 1260 (D.C. Cir. 1995) 18

Falk v. Gen. Motors Corp.,
496 F. Supp. 2d 1088 (N.D. Cal. 2007) 4

First Marblehead Corp. v. House,
473 F.3d 1 (1st Cir. 2006) 5

Frances T. v. Vill. Green Owners Ass'n,
42 Cal.3d 490 (1986) 6

Garrison v. State of Louisiana,
379 U.S. 64 (1964) 18

Gilmore v. Wells Fargo Bank N.A.,
75 F. Supp. 3d 1255 (N.D. Cal. Dec. 16, 2014) 5

Greater Houston Transp. Co. v. Uber Techs., Inc.,
155 F. Supp. 3d 670 (S.D. Tex. 2015) 12

Hardee's of Maumelle, Ark., Inc. v. Hardee's Food Sys., Inc.,
31 F.3d 573 (7th Cir. 1994) 17

Haskins v. Symantec Corp.,
654 F. App'x 338 (9th Cir. 2016) 16

In re Apple Inc. Sec. Litig.,
2020 WL 2857397 (N.D. Cal. June 2, 2020) 12, 18

In re Carrier IQ, Inc.,
78 F. Supp. 3d 1051 (N.D. Cal. 2015) 16

In re Chrysler-Dodge-Jeep Ecodiesel Mktg., Sales Pracs., & Prods. Liab. Litig.,
295 F. Supp. 3d 927 (N.D. Cal. 2018) 11, 15

In re Coventry Healthcare, Inc. Sec. Litig.,
2011 WL 1230998 (D. Md. Mar. 30, 2011) 11

In re Ford Motor Co. Sec. Litig.,
381 F. 3d 563 (6th Cir. 2004) 12

In re Fresenius Granuflo/Naturalyte Dialysate Prod. Liab. Litig.,
76 F. Supp. 3d 321 (D. Mass. 2015) 7

In re MyFord Touch Consumer Litig.,
46 F. Supp. 3d 936 (N.D. Cal. 2014) 15

In re Quality Sys., Inc. Sec. Litig.,
865 F.3d 1130 (9th Cir. 2017) 12

In re Takata Airbag Prods. Liability Litig.,
2017 WL 775811 (S.D. Fla. Feb. 27, 2017) 10

In re Volkswagen Timing Chain Prod. Liab. Litig.,
2017 WL 1902160 (D.N.J. May 8, 2017) 8

In re Yahoo! Inc. Customer Data Sec. Breach Litig.,
2017 WL 3727318, *29 (N.D. Cal. Aug. 30, 2017) 12

In re ZF-TRW Airbag Control Units Prods. Liab. Litig.,
601 F. Supp. 3d 625 (C.D. Cal. 2022) 16, 17

In re: Gen. Motors LLC Ignition Switch Litig.,
2016 WL 3920353 (S.D.N.Y. July 15, 2016) 10

Jones v. Corus Bankshares, Inc.,
701 F. Supp. 2d 1014 (N.D. Ill. 2010) 11

Kaloti Enter., Inc. v. Kellogg Sales Co.,
699 N.W.2d 205 (Wis. 2005) 8

Kearns v. Ford Motor Co.,
567 F.3d 1120 (9th Cir. 2009) 4

Khan v. Shiley Inc.,
266 Cal. Rptr. 106 (Cal. Ct. App. 1990) 10

Kottle v. Nw. Kidney Centers,
146 F.3d 1056 (9th Cir 1998) 18

Lerner v. DMB Realty, LLC,
234 P.3d 909 (Ariz. Ct. App. 2014) 8

LiMandri v. Judkins,
60 Cal. Rptr. 2d 539 (1997) 6, 10

Lloyd v. Facebook, Inc.,
2022 WL 4913347 (N.D. Cal. Oct. 3, 2022) 7

MacDonald v. Ford Motor Co.,
37 F. Supp. 3d 1087 (N.D. Cal. 2014) 5, 13, 16

Maxwell v. United Servs. Auto. Ass'n,
342 P.3d 474 (Colo. App. 2014) 13

McCabe v. Daimler AG,
948 F. Supp. 2d 1347 (N.D. Ga. 2013) 8

McDonald v. Smith,
472 U.S. 479 (1985) 18

McKee v. James,
2013 WL 3893430 (N.C. Super. July 24, 2013) 8

Mendiondo v. Centinela Hosp. Med. Ctr.,
521 F.3d 1097 (9th Cir. 2008) 4

Mewawalla v. Middleman,
601 F. Supp. 3d 574 (N.D. Cal. 2022) 6

Mirkin v. Wasserman,
5 Cal. 4th 1082 (Cal. 1993) 13

Morris v. Princess Cruises, Inc.,
236 F.3d 1061 (9th Cir. 2001) 12

Njoku v. Geico Gen. Ins. Co.,
2020 WL 4915433, (N.D. Cal. May 6, 2020). 5

Nooner Holdings, Ltd. v. Abilene Vill., LLC,
668 S.W.3d 956 (Tex. App. 2023) 10

Nota Const. Corp. v. Keyes Assocs., Inc.,
694 N.E.2d 401 (Mass. App. Ct. 1998) 9

O'Connor v. Uber Techs., Inc.,
2013 WL 6354534 (N.D. Cal. Dec. 5, 2013) 6, 7

Patino v. Cnty. of Monterey,
2023 WL 375349 (N.D. Cal. Jan. 24, 2023) 7

Perkins v. LinkedIn Corp.,
53 F. Supp. 3d 1190 (N.D. Cal. 2014) 16

Perry v. Merit Sys. Prot. Bd.,
582 U.S. 420 (2017) 18

Pirozzi v. Apple Inc.,
913 F. Supp. 2d 840 (N.D. Cal. 2012) 16

Priselac v. Chemours Co.,
2022 WL 909206 (E.D.N.C. Mar. 28, 2022) 6

Quashnock v. Frost,
445 A.2d 121 (Pa. Super. Ct. 1982) 8

Roberts v. Paine,
199 A. 112 (Conn. 1938) 9

Sloan v. Gen. Motors LLC,
287 F. Supp. 3d 840 (N.D. Cal. 2018) 5, 13, 16

Social Media Cases,
JCCP 5255, 2023 WL 6847378 (Cal. Super. Oct. 13, 2023) 10

Sosa v. DIRECTV, Inc.,
437 F.3d 923 (9th Cir. 2006) 18

Stamm v. Salomon,
551 S.E. 2d 152 (N.C. Ct. App. 2001) 10

Sunquest Info. Sys., Inc. v. Dean Witter Reynolds, Inc.,
40 F. Supp. 2d 644 (W.D. Pa. 1999) 10

Tabler v. Panera LLC,
2019 WL 5579529 (N.D. Cal. Oct. 29, 2019) 16

Theme Promotions, Inc. v. News Am Mktg. FSI,
546 F.3d 991 (9th Cir. 2008) 18

Tuosto v. Philip Morris USA Inc.,
2007 WL 2398507 (S.D.N.Y. Aug. 21, 2007) 18

TVT Recs. v. Island Def Jam Music Grp.,
412 F.3d 82 (2d Cir. 2005) 8

XYZ Two Way Radio Serv., Inc. v. Uber Techs., Inc.,
214 F. Supp. 3d 179 (E.D.N.Y. 2016) 12

Zwiercan v. Gen. Motors Corp.,
2002 WL 31053838 (Pa. Ct. Com. Pl. 2002) 13

Statutes

18 U.S.C. § 1001 18

Rules

Fed. R. Civ. P. 12(b)(6) 4

Fed. R. Civ. P. 9(b) 4, 5, 13, 16

Other Authorities

63A Am. Jur. 2d Products Liability § 779 10

Restatement (Second) of Torts (1965) 11

Stan Lee & Steve Ditko, *Introducing Spider Man*,
Amazing Fantasy 15 (Marvel Comics 1962) 2

U.S. Const. Amend. I 18

I. INTRODUCTION

Mark Zuckerberg is not merely a CEO; he is a household name. Given his outsized role as Meta's spokesperson, his superior knowledge of its products—namely, Facebook and Instagram—and his decision to vouch for their safety, Mr. Zuckerberg accepted the duty to speak fully and truthfully on the risks Meta's platforms pose to children's health. Had he done so, Plaintiffs would have acted differently to prevent the suffering they endured. Mr. Zuckerberg's omissions harmed Plaintiffs[1] and countless children across the country. He is an actual participant in the tort, and his efforts to hide behind Meta are unavailing. His motion to dismiss, ECF 518 (hereinafter "Mot."), should be denied.

II. BACKGROUND

A. Mark Zuckerberg's power as a decision-maker and spokesperson for Meta is unparalleled.

Mark Zuckerberg delivered Facebook into existence, spearheaded his company's acquisition of Instagram, and shepherded explosive growth on both platforms, making Facebook and Instagram ubiquitous in public life. SAC[2] ¶¶ 187, 193-208, 209-15. Even at Facebook's inception, Mr. Zuckerberg knew he was no ordinary "Founder," listing himself also as the "Master and Commander" and "Enemy of the State" on Facebook's first masthead.[3] His unparalleled role in the company persists to this day, *see* AG ¶ 39,[4] and issues he does not engage with do not get institutional support. Mr. Zuckerberg repeatedly ignored employee requests to invest in initiatives addressing user well-being, leaving Meta's mental health team defunded such that it "completely stopped" work by the end of 2019. SAC ¶ 366. In 2021, a renewed request to address "concerns about the impact of [Meta's] products on young

---

[1] Mr. Zuckerberg notes that only the twenty-four Plaintiffs who have filed short form complaints naming him as a defendant are the subject of his motion. Mot. at 2. The twenty-five Plaintiffs named in the caption represent the updated and active cases in which Mr. Zuckerberg is named as a defendant. Each of these Plaintiffs used Instagram.

[2] References to "SAC ¶ __" are to Plaintiffs' Second Amended Master Complaint (Personal Injury), ECF 494.

[3] *See* Jillian D'Onfro, *Facebook's News Feed is 10 years old. This is how the site has changed*, World Econ. F. (Sept. 9, 2016), https://weforum.org/agenda/2016/09/facebooks-news-feed-is-10-years-old-this-is-how-the-site-has-changed (referenced in SAC ¶ 195 nn.209, 210, 211, 213).

[4] References to "AG ¶ __" are to the paragraphs of the multistate complaint filed by numerous attorneys general in *State of Arizona v. Meta Platforms, Inc.*, No. 4:23-cv-05448, ECF 1 (Oct. 24, 2023), which are incorporated by reference in SAC ¶ 391A.

people's mental health," was met with a Meta executive's acknowledgment that there was a "very low-likelihood that Mark chooses to fund more here." AG ¶¶ 621-29.

With great power comes great responsibility.[5] Unfortunately, Mr. Zuckerberg has not lived up to that maxim. Even amidst Meta's unprecedented evolution and expansion, he has maintained a tight grip on design and engagement-focused decisions, consistently promoting growth and profits over user safety. *See, e.g.*, AG ¶¶ 57, 144. As one Meta software engineer explained, "many employees feel that if they whistleblow, dissent, give feedback to unethical decisions, etc, then they are at risk for being fired." SAC ¶ 381. When Meta executives were in active disagreement about whether to increase the quantity of notifications sent to users to boost engagement, the conversation ended with confirmation that the number of active users "[was] a bigger concern for Mark . . . than user experience." AG ¶ 322. Another such veto included rejecting his Vice President of Product Design and Responsible Innovation's proposed ban on filters that simulate plastic surgery. AG ¶¶ 339-63. Although "outside academics and experts consulted were nearly unanimous on the harm," and the proposed ban earned significant employee support, Mr. Zuckerberg ignored the data, called the proposal "paternalistic," and cited "clear[] demand" as reason enough to reject it. *Id.* ¶¶ 339-63.

B. Mark Zuckerberg concealed evidence that Meta's platforms are not safe for youth.

Mr. Zuckerberg has long established himself as a resource to the public on all things Meta—he is *the* face of Facebook, and now Meta. For example, in 2006, when users expressed concern over the novel Facebook "news feed" displaying more data about users in real time, Mr. Zuckerberg personally responded. *See* SAC ¶ 195 n.213 (citing article referencing Zuckerberg post titled "Calm down. Breathe. We hear you."). In the wake of outcry about extremist content on social media, Mr. Zuckerberg posted a lengthy "Blueprint for Content Governance and Enforcement." SAC ¶ 271. People listen to what Mr. Zuckerberg has to say. And Mr. Zuckerberg's ability to command attention and drive the news is not limited to his own products. He publishes op-eds in major papers, is a guest of prominent technology journalists, and is invited to public conversations with senators. *See* SAC ¶ 369 nn.490, 497, 484. His every word is a focus of such intense public interest that Marquette University hosts The

[5] Stan Lee & Steve Ditko, *Introducing Spider Man*, Amazing Fantasy 15 (Marvel Comics 1962).

Zuckerberg Files, a digital archive of Mr. Zuckerberg's public statements. SAC ¶ 176 n.167 (citing same).

Mr. Zuckerberg has long understood his influence over how people understand and use his products. Regrettably, he abused that power by concealing information about Facebook and Instagram and creating a false impression of their safety. Mr. Zuckerberg was warned by both external and internal sources that Meta's products are unsafe for children. In 2019, he met with a psychologist and leading expert who explained that research "points heavily to a connection" between social media use and youth mental health issues. SAC ¶ 365. Likewise, a Meta employee informed him that the company was "not on track to succeed for our core well-being topics (problematic use, bullying & harassment, connections, and [suicide and self-injury]), and [was] at increased regulatory risk and external criticism. These affect everyone, especially Youth and Creators." SAC ¶ 183. Mr. Zuckerberg received a similar warning in a 2021 email from an employee, who cautioned that users were having far more harmful experiences on Instagram than Meta's public metrics acknowledged. AG ¶¶ 504-505. On top of this, Mr. Zuckerberg knew that children under the age of 13 were using Meta products. Internal Meta documents indicate that in 2018, Mr. Zuckerberg personally received a report estimating there were four million kids under thirteen using Instagram. AG ¶¶ 657-60.

Mr. Zuckerberg easily could have, but chose not to, use his megaphone to share any of this information with the public. Instead, he did the opposite. In 2019, as he ignored internal requests to fund user well-being initiatives, Mr. Zuckerberg posted to his followers: "You should expect we'll do everything we can to keep you safe on our services," SAC ¶ 370(l), and told investors he was "proud of the work that [Meta] ha[s] done to get in front of a lot more of these [safety and security] issues." SAC ¶ 370(j). And in 2021, despite knowing that Meta's algorithms could contribute to addiction (or what Meta called "problematic use"), SAC ¶¶ 276-79, Mr. Zuckerberg told Congress—and the American public—that he did not believe his platforms harm children. SAC ¶ 370(p). At the same hearing, when asked whether passive consumption of social media harmed children's mental health—something Meta's internal research established—Mr. Zuckerberg suggested the opposite was true: "Overall, the research that we have seen is that using social apps to connect with other people can have positive mental health benefits and well-being benefits by helping people feel more connected and less lonely." AG ¶ 419. And when two United States senators wrote to Mr.Zuckerberg asking for the findings of

Meta's research on the platforms' impact on youth well-being, the response did not disclose Meta's own detailed studies demonstrating its products can and do result in "problematic use" that disproportionately causes harm for young users. AG ¶ 593; SAC ¶¶ 373-79. The truth became front-page news[6] just weeks later when Frances Haugen shared these studies and other Meta documents with the Wall Street Journal and testified before Congress. SAC ¶¶ 217-20, 377.

Furthermore, since at least 2011, Mr. Zuckerberg has maintained that Meta does not allow children under the age of thirteen to use its products. SAC ¶¶ 369(c). He reiterated this point at a congressional hearing in 2021, stating: "[W]e have additional systems that try to determine what someone's age might be so if we detect that someone might be under the age of 13, even if they lied, we kicked them off." AG ¶ 770. Mr. Zuckerberg's incomplete and misleading public statements created a false impression that Meta was actively working to make its products safe for minors and inaccessible to children too young for the platforms.

III. LEGAL STANDARDS

"Dismissal under Rule 12(b)(6) is appropriate only where the complaint lacks a cognizable theory or sufficient facts to support a cognizable legal theory." *Mendiondo v. Centinela Hosp. Med. Ctr.*, 521 F.3d 1097, 1104 (9th Cir. 2008). Allegations of "factual content that allows the court to draw the reasonable inference that the defendant is liable for the misconduct alleged" suffice to survive a 12(b)(6) motion. *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009). The court must accept factual allegations in the complaint as true in deciding the claims' plausibility. *Id.*

While the pleading requirements for claims that "sound in fraud" or are "grounded in fraud," are heightened, Fed. R. Civ. P. 9(b); *Kearns v. Ford Motor Co.*, 567 F.3d 1120, 1125 (9th Cir. 2009), "many courts do not apply a strict Rule 9(b) analysis to allegations of omissions." *Belville v. Ford Motor Co.*, 60 F. Supp. 3d 690, 696 (S.D.W. Va. 2014); *see also Falk v. Gen. Motors Corp.*, 496 F. Supp. 2d 1088, 1098-99 (N.D. Cal. 2007) (omission claim did not have to "specify the time, place, and specific content of an omission as precisely as would a . . . false representation claim"). "Because the Plaintiffs are alleging a

[6] The print version of at least two articles included in the "Facebook Files" cited in paragraph 217 of the SAC appeared on page A1. John D. McKinnon & Ryan Tracy, *Facebook Hearing Fuels Call for Reins on Tech*, Wall St. J., October 6, 2021, at A1; Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Platform Healthier. It Got Angrier Instead*, Wall St. J., September 16, 2021, at A1.

failure to act instead of an affirmative act, the Plaintiffs cannot point out the specific moment when the Defendant failed to act." *MacDonald v. Ford Motor Co.*, 37 F. Supp. 3d 1087, 1096 (N.D. Cal. 2014) (cleaned up). Therefore, an omissions-based fraud claim can satisfy Rule 9(b) through more generalized allegations about the "who what when where, and how" of the alleged misconduct: who should have revealed information, what the information was, when they should have done so, and where plaintiffs would have received this information. *Id.*

IV. ARGUMENT

Plaintiffs assert fraudulent- and negligent-omission claims against both Meta and its "Master and Commander." As discussed in more detail below, together, the Second Amended Master Complaint and applicable short-form complaints more than adequately allege that Mr. Zuckerberg (1) owed a duty to disclose material facts to the Plaintiffs and (2) "concealed or suppressed" those facts, and that (3) Plaintiffs were unaware of the facts and would have acted differently if they had known the truth, which (4) resulted in their injury.[7] *See, e.g.*, *Sloan v. Gen. Motors LLC*, 287 F. Supp. 3d 840, 865 (N.D. Cal. 2018) (quoting *Hahn v. Mirda*, 54 Cal. Rptr. 3d 527, 32 (Cal. Ct. Ap. 2007)).

Mr. Zuckerberg moves to dismiss on the grounds that Plaintiffs have failed to allege reliance, injury, or actionable misstatements of fact within the framework of fraudulent misrepresentation.[8] But Plaintiffs' claims center on what Mr. Zuckerberg *failed* to say. In this regard, Plaintiffs' claims are largely untouched by Mr. Zuckerberg's arguments and authorities, which focus on the actionability of affirmative false statements, rather than misrepresentation by omission. In the only portion of his

---

[7] As Mr. Zuckerberg notes, the fundamental elements of fraud are substantially similar from state to state. Mot. at 7. Negligent misrepresentation shares these elements and requires that reliance be "justifiable" and the omission made "with failure to exercise reasonable care or competence." *See, e.g.*, *First Marblehead Corp. v. House*, 473 F.3d 1, 9-10 (1st Cir. 2006).

[8] With scant exception, Mr. Zuckerberg only addresses Plaintiffs' negligent-misrepresentation claims indirectly, insofar as their elements overlap with fraudulent misrepresentation. *See* Mot. at 7 n.6 (citing one case noting the "reasonable reliance" standard applicable in negligent-misrepresentation cases). Accordingly, Plaintiffs do not separately discuss their negligence theory but respectfully submit that the facts discussed in part IV.B.3 demonstrate that their reliance was reasonable, even under the heightened pleading standard of Rule 9(b). However, Plaintiffs note that "[t]he Ninth Circuit has not yet decided whether Rule 9(b)'s heightened pleading standard applies to a claim for negligent misrepresentation," *Gilmore v. Wells Fargo Bank N.A.*, 75 F. Supp. 3d 1255, 1269 (N.D. Cal. Dec. 16, 2014), and district courts remain "divided on [this] question." *Njoku v. Geico Gen. Ins. Co.*, 2020 WL 4915433, *3 (N.D. Cal. May 6, 2020).

motion that squarely addresses Plaintiffs' omissions claims, Mr. Zuckerberg argues that he owes no duty to disclose the dangers Meta's products present to minors. As explained further in Part IV.B.1, however, Mr. Zuckerberg's duty to disclose arose from his (1) superior and exclusive knowledge about, and (2) misleading partial representations regarding, the risk of harm created by his company's products. Mr. Zuckerberg's motion should be denied.

A. Mark Zuckerberg is personally liable for his omissions and concealment.

As an initial matter, Mr. Zuckerberg insists that he is insulated from liability for Meta's conduct. Mot. at 4-6. In doing so, he mischaracterizes Plaintiffs' claims as allegations based on Mr. Zuckerberg's "[m]ere status as an executive" or "shareholder" *Id.* at 5. Plaintiffs are not suing Mr. Zuckerberg for Meta's conduct—they are suing him for *his own* tortious omissions and concealment.

Mr. Zuckerberg may not "hide 'behind the corporation where he is an actual participant in the tort.'" *Coastal Abstract Serv., Inc. v. First Am. Title Ins. Co.*, 173 F.3d 725, 734 (9th Cir. 1999) (quoting *Donsco, Inc. v. Casper Corp.*, 587 F.2d 602, 606 (3d Cir. 1978)); *see also O'Connor v. Uber Techs., Inc.*, 2013 WL 6354534, at *18 (N.D. Cal. Dec. 5, 2013) (quoting *Frances T. v. Vill. Green Owners Ass'n*, 42 Cal.3d 490, 507-08 (1986)) ("[T]he corporate fiction . . . was never intended to insulate officers from liability from their own tortious conduct."); *Priselac v. Chemours Co.*, 2022 WL 909206, at *11 (E.D.N.C. Mar. 28, 2022) (plaintiff sufficiently alleged two corporate officers' participation in environmental trespass claim by alleging omitted material information in permit applications they helped craft). Indeed, the case Mr. Zuckerberg cites as representative of the black-letter law on corporate-officer liability supports denying his motion. Mot. at 4-5. As in *Mewawalla v. Middleman*, 601 F. Supp. 3d 574 (N.D. Cal. 2022), Plaintiffs have alleged sufficient facts to show "that [Mr. Zuckerberg] had a duty to disclose," *infra* IV.B.1, that Mr. Zuckerberg "had exclusive knowledge of material facts not known to the plaintiff[s]," and "to put [Mr. Zuckerberg] on notice of what [he] concealed from Plaintiff[s] and when it occurred." *Mewawalla*, 601 F. Supp. 3d at 602-03 (quoting *LiMandri v. Judkins*, 60 Cal. Rptr. 2d 539 (1997)).

Far from "rely[ing] on allegations regarding Meta's corporate conduct generally," Mot. at 6, Plaintiffs allege that Mr. Zuckerberg personally omitted and concealed material facts about his company. *See generally supra* Part II.B. Plaintiffs detail Mr. Zuckerberg's actual and constructive knowledge of the dangers Meta's products pose to the mental health of minors, *see* SAC ¶¶ 183, 276-

81, 365; AG ¶¶ 267, 521, 613, 621-29, his role in creating those dangers, *see* AG ¶¶ 340-58, the widespread use of the products by children under 13-years-old, AG ¶¶ 657-60, and various invitations Mr. Zuckerberg had to share this information with the public, *see, e.g.*, SAC ¶ 370(p) (asking Mr. Zuckerberg, "Do you believe that your platform harms children?"); AG ¶ 419 (asking Mr. Zuckerberg whether passively consuming social media harms children's mental health); SAC ¶¶ 373 n.508, 374-79, AG ¶ 593 (asking Mr. Zuckerberg whether Meta "conducted research on the effect of its platforms and products on children's and teens' mental health and well being" and the findings of such studies); SAC ¶¶ 369(c) (asking Mr. Zuckerberg to "elaborate upon" Facebook's accessibility to children under thirteen).

Mr. Zuckerberg also argues that because claims against him in unrelated cases were dismissed in the past, Plaintiffs' claims, too, are ripe for dismissal. Mot. at 5. Again, Mr. Zuckerberg fails to take the claims against him in this case seriously. They are a far cry from merely "identifying [his] role[] in the corporation and alleging that [he was] 'responsible' for" existing corporate policies or practices. *O'Connor*, 2013 WL 6354534, at *18; *see also Lloyd v. Facebook, Inc.*, 2022 WL 4913347, at *5 (N.D. Cal. Oct. 3, 2022) (complaint alleged only the bare assertion that Mr. Zuckerberg "should be held liable because he is the CEO of the company and was 'personally involved and/or directed the challenged acts'" with no additional facts); *Brock v. Zuckerberg*, 2021 WL 2650070, at *4 (S.D.N.Y. June 25, 2021) ("Plaintiff fails to allege any facts that connect Zuckerberg . . . to Facebook's removal of his posts").[9]

B. Plaintiffs plausibly allege Mark Zuckerberg's fraudulent omissions and concealment.

1. *Mr. Zuckerberg had a duty to disclose the dangers Meta's products present to minors.*

Mr. Zuckerberg had a duty to disclose the risks Meta's products pose to children's mental health and physical well-being. Plaintiffs plausibly allege facts giving rise to such a duty based upon:

[9] Mr. Zuckerberg's other cited authorities are equally distinguishable. *See Patino v. Cnty. of Monterey*, 2023 WL 375349, at *4 (N.D. Cal. Jan. 24, 2023) (complaint "lump[ed] multiple defendants together in a manner that ma[de] it impossible to tell what each defendant is alleged to have done or not done"); *In re Fresenius Granuflo/Naturalyte Dialysate Prod. Liab. Litig.*, 76 F. Supp. 3d 321, 336 (D. Mass. 2015) (the plaintiff "failed to produce any evidence suggesting the existence of red flags that were brought to [Defendant's] attention" after discovery).

(1) Mr. Zuckerberg's exclusive and superior knowledge of the ways Meta's products exploit minors, presenting risks to their health; and (2) his public, partial representations concerning the safety of Meta's products.

Courts overwhelmingly recognize that the duty to disclose arises "where one party possesses superior knowledge, not readily available to the other, and knows that the other is acting on the basis of mistaken knowledge." *TVT Recs. v. Island Def Jam Music Grp.*, 412 F.3d 82, 91 (2d Cir. 2005); *see, e.g.*, *Bank of Montreal v. Signet Bank*, 193 F.3d 818, 829 (4th Cir. 1999) (applying Virginia law) ("A duty may arise . . . if the fact is material and the one concealing has superior knowledge and knows the other is acting upon the assumption that the fact does not exist.").[10]

In the present case, Plaintiffs sufficiently plead Mr. Zuckerberg's superior knowledge of the harms Meta's products pose to minors. *See, e.g.*, *In re Volkswagen Timing Chain Prod. Liab. Litig.*, 2017 WL 1902160, at *19 (D.N.J. May 8, 2017). Plaintiffs allege that Mr. Zuckerberg "heard firsthand from a leading researcher that Instagram and Facebook posed unique dangers to young people," SAC ¶¶ 185,

---

[10] *See also Coldwell Banker Whiteside Assocs. v. Ryan Equity Partners, Ltd.*, 181 S.W.3d 879, 888 (Tex. App. 2006) ("The duty to disclose arises when one party knows that the other party is ignorant of the true facts and does not have an equal opportunity to discover the truth." (citations omitted)); *Lerner v. DMB Realty, LLC*, 234 P.3d 909, 917 (Ariz. Ct. App. 2014) (a party to a transaction "may be required to disclose information when the [other party] reasonably cannot discover the information for himself"); *BAC Home Loans Serv. v. Farina*, 2010 Conn. Super. LEXIS 4929, *2 (Conn. Super. Ct. June 2, 2010) ("[A] duty to disclose arises where one party's superior knowledge of essential facts renders a transaction without disclosure inherently unfair."); *Dean v. Beckley*, 2010 WL 3928650, at *5 (D. Md. Oct. 1, 2010) (acknowledging a duty to disclose "if the fact is material and the one concealing has superior knowledge and knows the other is acting upon the assumption that the fact does not exist"); *McKee v. James*, 2013 WL 3893430, at *8 (N.C. Super. July 24, 2013) (Recognizing a duty to disclose "where one party has knowledge of a latent defect in the subject matter of the negotiations about which the other party is both ignorant and unable to discover through reasonable diligence." (citation omitted)); *Quashnock v. Frost*, 445 A.2d 121, 128 (Pa. Super. Ct. 1982) (party with superior knowledge had a duty to disclose a latent or "serious and dangerous condition . . . not readily observable upon reasonable inspection."); *First Choice Armor & Equip., Inc. v. Toyobo Am., Inc.*, 717 F. Supp. 2d 156, 162 (D. Mass. 2010) (defendant's "position of 'superior knowledge' with respect to plaintiff . . . triggered a duty to disclose"); *Bain v. Jackson*, 783 F. Supp. 2d 13, 18 (D.D.C. 2010) (applying New York law) (a "duty to disclose arises only where one party possesses superior knowledge of essential facts that makes a transaction inherently unfair"); *McCabe v. Daimler AG*, 948 F. Supp. 2d 1347, 1368 (N.D. Ga. 2013) (acknowledging duty to disclose where defendant concealed "intrinsic qualities of the article which the other party by the exercise of ordinary prudence and caution could not discover" (citation omitted)); *Kaloti Enter., Inc. v. Kellogg Sales Co.*, 699 N.W.2d 205, 213 (Wis. 2005) (establishing duty to disclose where a material fact "is peculiarly and exclusively within the knowledge of one party, and the mistaken party could not reasonably be expected to discover it").

365, 993, "was warned personally" by employees that Meta was "not on track to succeed for [its] core well-being topics . . . affect[ing] everyone, especially Youth," *id.* ¶ 183, "came to understand that Meta was 'actively encouraging young girls into body dysmorphia,'" AG ¶¶ 340-41, received data from internal researchers confirming that social media exacerbates negative comparisons and user mental health issues, *id.* ¶¶ 557-58, 416-17, and declined to fund internal initiatives to develop solutions to mounting "concerns about the impact of [Meta's] products on young people's mental health," *id.* ¶¶ 612-29. Plaintiffs allege sufficient facts to show that Mr. Zuckerberg had more knowledge (from more sources) about the harms Meta's products pose to youths than anyone else, and that he repeatedly concealed it. Given Mr. Zuckerberg's prominence as the speaker for a pioneering technology used by billions of people, he owed a duty to provide Plaintiffs with material information about the dangers Meta's platforms pose to minors, about which he had exclusive and far superior knowledge.

Mr. Zuckerberg is incorrect that a "a duty to disclose [typically] does not exist absent a contractual or other special relationship between the parties." Mot. at 12. Although the duty-to-disclose inquiry often begins by evaluating the "relationship of the parties," the issue boils down to "whether the occasion and circumstances are such as to impose a duty to speak." *Roberts v. Paine*, 199 A. 112, 115 (Conn. 1938). In lieu of artificially restricting application of the duty to disclose to specific transactions and business relationships, courts favor a contextual approach. *See, e.g.*, *Nota Const. Corp. v. Keyes Assocs., Inc.*, 694 N.E.2d 401, 404 (Mass. App. Ct. 1998) ("[A] duty to disclose may arise in a number of circumstances"). In general, "a person has a duty to disclose to another with whom he deals facts that 'in equity or good conscience' should be disclosed," even where no contractual or fiduciary relationship exists. *Berger v. Sec. Pac. Info. Sys., Inc.*, 795 P.2d 1380, 1383 (Colo. App. 1990) (employer had a duty to disclose information to prospective employee).

With these points in mind, Mr. Zuckerberg's duty must be considered in the context of his company's. In the process of founding and leading a company that revolutionized social engagement and communication—and captured the attention of millions of people across the globe—Mr. Zuckerberg was the trusted voice on all things Meta, including on sensitive, consequential social and political issues. *See supra* Part II.B. Even as he commanded an increasingly towering presence in the tech industry, Mr. Zuckerberg remained an approachable resource to the public. *See, e.g.*, SAC ¶ 195

n.213 ("Calm down. Breathe. We hear you."). Plainly, Meta owed a duty to its customers to inform them that Meta's products are defective and to warn them of the risks associated with their use.[11] *Social Media Cases*, JCCP 5255, 2023 WL 6847378, at *44-45 (Cal. Super. Oct. 13, 2023); *see also Khan v. Shiley Inc.*, 266 Cal. Rptr. 106, 112 (Cal. Ct. App. 1990) (confirming that manufacturers have a duty to disclose safety information concerning anything from a "mechanical heart valve [to] frozen yogurt"); *In re: Gen. Motors LLC Ignition Switch Litig.*, 2016 WL 3920353, at *39 (S.D.N.Y. July 15, 2016) (the plaintiff adequately plead violation of duty to disclose where defendant knew about latent defects). By cultivating his roles in public life as both the embodiment of Meta and Silicon Valley's approximation of a philosopher king, Mr. Zuckerberg accepted the same duty.

Mr. Zuckerberg also had a duty to disclose his knowledge of the harms social media posed to young people because of his partial representations on the subject. "[W]hen the defendant makes partial representations but also suppresses some material facts," a duty to disclose exists, 63A Am. Jur. 2d Products Liability § 779,[12] and "there may be recovery either on the basis of the original misleading

---

[11] In its motion to dismiss the Personal Injury Plaintiffs' misrepresentation claims (SAC Counts 8 and 9), Meta does not argue that it did not have a duty to disclose its platforms' defects to users. ECF 517 at 64-66.

[12] *See, e.g.*, *In re Takata Airbag Prods. Liability Litig.*, 2017 WL 775811, at *4 (S.D. Fla. Feb. 27, 2017) (applying South Carolina Law) (finding duty to disclose where defendant made "incomplete representations about the safety and reliability of [products]"); *Bays v. Hunter Sav. Ass'n*, 539 F. Supp. 1020, 1025 (S.D. Ohio 1982) ("Ohio law imposes a duty to make full disclosure in circumstances where full disclosure is necessary to dispel misleading impressions that are, or might have been, created by partial revelation of the facts."); *Stamm v. Salomon*, 551 S.E. 2d 152, 158 (N.C. Ct. App. 2001) (concealment occurs when a person "has made a partial or incomplete representation"); *Nooner Holdings, Ltd. v. Abilene Vill., LLC*, 668 S.W.3d 956, 966 (Tex. App. 2023) ("Various Texas courts of appeals have generally agreed that a duty to disclose may arise based upon a partial disclosure . . . [and] when a party makes a true disclosure that conveys a false impression."); *LiMandri v. Judkins*, 52 Cal.App.4th 326, 336, 60 Cal.Rptr.2d 539 (1997) (defendant has a duty to disclose "when [he] makes partial representations but also suppresses some material facts"); *Sunquest Info. Sys., Inc. v. Dean Witter Reynolds, Inc.*, 40 F. Supp. 2d 644, 657 (W.D. Pa. 1999) ("a party making an 'incomplete' representation that could be misleading if left to stand alone is under a duty to disclose such other facts as may be necessary to make the initial statement clear"); *Burman v. Richmond Homes Ltd.*, 821 P.2d 913, 918 (Colo. App. 1991) ("[A] party has a duty to disclose if he has stated facts that he knows will create a false impression unless other facts are disclosed."); *Bank of Montreal v. Signet Bank*, 193 F.3d 818, 829 (4th Cir. 1999) (applying Virginia law) ("A duty may arise if . . . one party takes actions which divert the other party from making prudent investigations (e.g., by making a partial disclosure)"); *DiMichele v. Perrella*, 120 A.3d 551, 554 (Conn. App. Ct. 2015) ("Under the common law, a duty to disclose 'is imposed on a party insofar as he voluntarily makes disclosure. A party who assumes to speak must make a full and fair disclosure as to the matters about which he assumes to speak.'").

statement *or of the nondisclosure of the additional facts*." Restatement (Second) of Torts § 551 cmt. g (1965) (emphasis added). Moreover, "concealment may amount to fraud . . . where, in addition to a party's silence, there is *any statement, word, or act on his part*, which tends affirmatively to the suppression of the truth, or to a covering up or disguising of the truth, or to a withdrawal or distraction of a party's attention from the real facts." *In re Chrysler-Dodge-Jeep Ecodiesel Mktg., Sales Pracs., & Prods. Liab. Litig.*, 295 F. Supp. 3d 927, 1008-09 (N.D. Cal. 2018) (emphasis added) (quoting *Lubore v. RPM Assocs.*, 674 A.2d 547, 556 (Md. Ct. App. 1996)); *see also id.* at 1010 ("[i]f in addition to the party's silence there is any statement, even in word or act on his part, which tends affirmatively to a suppression of truth . . . the concealment becomes fraudulent" (internal quotation marks omitted)).

Here, Plaintiffs adequately allege that Mr. Zuckerberg made numerous statements in which he concealed material information about the risks of using Meta's platforms despite various explicit invitations to disclose it. *See supra* Part IV.A. For example, when Mr. Zuckerberg was asked about the harm that passive consumption of social media could pose to children's mental health, he "played up the benefits" of Meta's platforms despite "being given talking points on the negative effects of passive consumption on mental health to prepare for the congressional hearing." AG ¶ 419. In sum, Mr. Zuckerberg had a duty to disclose what he knows better than anyone: that using Meta's products could harm the mental health of children.

2. *Plaintiffs adequately plead that Mr. Zuckerberg made incomplete statements of fact.*

Mr. Zuckerberg's insistence that his partial disclosures regarding the safety of Meta's platforms constitute non-actionable "puffery" is wrong. Mot. at 9-11. Even a statement that could be construed as mere puffery is rendered materially misleading—giving rise to a duty to disclose—where the defendant's knowledge renders the statement deceptive. *See Jones v. Corus Bankshares, Inc.*, 701 F. Supp. 2d 1014, 1027-28 (N.D. Ill. 2010) ("[S]tatements that would otherwise amount to puffery can be actionable if the speaker is aware that the statement is deceptive."); *In re Coventry Healthcare, Inc. Sec. Litig.*, 2011 WL 1230998, at *12 (D. Md. Mar. 30, 2011).

Here, plaintiffs adequately allege that Mr. Zuckerberg made statements over several years indicating that Meta prioritized safety on its platforms and that the platforms are in fact safe, with

knowledge that those statements were incomplete and misleading. *See supra* Part II.B; *see also In re Apple Inc. Sec. Litig.*, 2020 WL 2857397, at *15 (N.D. Cal. June 2, 2020) ("[A] party cannot affirmatively create a positive impression of an area it knows to be doing poorly."); *In re Quality Sys., Inc. Sec. Litig.*, 865 F.3d 1130, 1143 (9th Cir. 2017) ("Even 'general statements of optimism, when taken in context, may form a basis for a securities fraud claim' when those statements address specific aspects of a company's operation that the speaker knows to be performing poorly.").

The authorities Mr. Zuckerberg cites in support of finding his statements non-actionable are inapposite. Mot. at 9-10. In two cases, the defendant made statements regarding safety without the existence of contrary facts at the time the statement was made. *In re Ford Motor Co. Sec. Litig.*, 381 F. 3d 563, 570 (6th Cir. 2004) (Ford disclosed "accurate historical data" without noting potential for less favorable results in future quarters); *Azoulai v. BMW of N. Am. LLC*, 2017 WL 1354781, at * 8 (N.D. Cal. Apr. 13, 2017) (under statutory claims asserted under omission theory, plaintiffs failed to assert defect contrary to any affirmative statements). In others, the plaintiffs did not assert any omission or concealment claim. *Morris v. Princess Cruises, Inc.*, 236 F.3d 1061 (9th Cir. 2001); *Greater Houston Transp. Co. v. Uber Techs., Inc.*, 155 F. Supp. 3d 670, 682 (S.D. Tex. 2015); *XYZ Two Way Radio Serv., Inc. v. Uber Techs., Inc.*, 214 F. Supp. 3d 179, 184 (E.D.N.Y. 2016). And the fraudulent-omission claim analysis in *In re Yahoo!* assumed a duty to disclose to dismiss on other grounds. *In re Yahoo! Inc. Customer Data Sec. Breach Litig.*, 2017 WL 3727318, *29 (N.D. Cal. Aug. 30, 2017). In light of Plaintiffs' factual allegations, Mr. Zuckerberg's attempt to recast his misleading statements (which give rise to a duty to disclose) as "involv[ing] vague, non-specific matters of ***opinion***," Mot. 10 (emphasis in original); Appx. A, misses its mark.

3. *Plaintiffs relied on and were harmed by Mr. Zuckerberg's misleading statements.*

In arguing that Plaintiffs did not adequately allege reliance, Mot. at 7-9, Mr. Zuckerberg again fails to appreciate the distinction between fraudulent misrepresentation and fraudulent omission and concealment. Plaintiffs' claims turn on the impact of Mr. Zuckerberg's *silence*—of his decision to hide highly material information about the safety of Instagram and Facebook from public view, despite speaking about the same topics. Thus, it does not matter whether Plaintiffs "saw any of the supposedly

false statements" or relied on any one specific statement by Mr. Zuckerberg, Mot. at 7-8. "What matters in an omission case . . . is whether the plaintiff had an opportunity to receive and therefore rely on the omitted information, not whether they actually received some other, irrelevant information." *Sloan v. Gen. Motors LLC*, 287 F. Supp. 3d 840, 875 (N.D. Cal. 2018). Under Rule 9(b), claims based on an omission can succeed without the same level of specificity required by a normal fraud claim. This is because a plaintiff alleging an omission-based fraud will not be able to specify the time, place, and specific content of an omission as would a plaintiff in a false representation claim. *See MacDonald*, 37 F. Supp. at 1096.

In this regard, the reliance inquiry focuses on whether plaintiffs would have acted differently had they known the disclosed information. In many states, a presumption of reliance arises when the omission is material. *See, e.g.*, *Cope v. Metro Life Ins. Co.*, 82 Ohio St.3d 426, 436 (Ohio 1998) ("It is not necessary to establish inducement and reliance upon material omissions by direct evidence. When there is nondisclosure of a material fact, courts permit inferences or presumptions of inducement and reliance."); *Zwiercan v. Gen. Motors Corp.*, 2002 WL 31053838, *5 (Pa. Ct. Com. Pl. 2002) ("Plaintiff's reliance on Defendant's active concealment can be presumed because the concealment of the alleged defect would be material to the Plaintiff's decision to purchase her car."). In others, "reliance may be inferred from circumstantial evidence where the defendant concealed a material fact from the plaintiff." *BP Am. Prod. Co. v. Patterson*, 263 P.3d 103, 110 (Colo. 2011); *see also Maxwell v. United Servs. Auto. Ass'n*, 342 P.3d 474, 481 (Colo. App. 2014) (explaining that the materiality of an omission allows for the inference of reliance through circumstantial rather than direct evidence, although there must still be some evidence of reliance). For example, in California, "[t]hat one would have behaved differently can be presumed, or at least inferred, when the omission is material," although plaintiffs must still plausibly allege that they would have been aware of the omitted information had it been publicly revealed. *Daniel v. Ford Motor Co.*, 806 F.3d 1217, 1225 (9th Cir. 2015); *see also Mirkin v. Wasserman*, 5 Cal. 4th 1082, 1093 (Cal. 1993) (explaining that materiality of omission is relevant to plaintiffs' allegations that, "had the omitted information been disclosed, [they] would have been aware of it and behaved differently").

Reading the Master Complaint and the Short Form Complaints together[13] and accepting as true all factual allegations in the pleadings, Plaintiffs adequately allege that they relied upon Mr. Zuckerberg's omissions. If Mr. Zuckerberg had disclosed how harmful Meta's platforms are to children, that material information would have been widely reported, and Plaintiffs both would have known and would have taken specific actions to prevent the harm that they and their families suffered.

For example, the *Baker* complaint contains allegations, omitted in Mr. Zuckerberg's selective quotation, Mot. at 8, that had he revealed the omitted information, "Plaintiff reasonably would have utilized more caution in their social media consumption, understood their symptoms were caused by social media sooner, and taken the necessary steps to mitigate damage to their mental health." *Baker* SFC, Ex. A (4:23-cv-01578). Other Plaintiffs and their parents allege that they would not have allowed their child to use Meta's platforms at all, *see e.g.*, *Booker* SFC (4:23-cv-01537) ("Had the omitted information been disclosed, Plaintiff Richard Neal Booker reasonably would have prohibited his minor child S.B. from ever downloading and using Instagram."); "would have delayed the age at which they got Plaintiff a phone until at least thirteen years old and would have looked for a phone with parental control features to control and monitor app downloads," *C.G.* SFC (4:23-cv-01568); and "would have taken action to prevent Plaintiff from accessing Instagram for many more years," *C.S.* SFC (4:23-cv-01569). Plaintiffs' specific averments of actions they would have taken had Mr. Zuckerberg not concealed information about the safety of Instagram demonstrate their reliance on Mr. Zuckerberg's omissions. *See McCabe*, 948 F. Supp. 2d at 1368 ("McCabe and Herring also allege that they expected to receive vehicles free from design or manufacturing defects and that they would not have purchased their vehicles had they known of the defect. . . . Thus, they have plausibly alleged justifiable reliance.").

Plaintiffs' individual allegations are further substantiated by the actions of Meta and its employees. Damningly, Meta executives who had full knowledge of the dangers Meta's platforms pose to children chose to ban or significantly reduce their children's use of social media, a choice Mr. Zuckerberg's concealment denied Plaintiffs and their loved ones. SAC ¶ 261. Meta—and

[13] Mr. Zuckerberg's attempt to isolate Plaintiffs' Short-Form Complaints from the allegations of the Master Complaint is not well taken. Plaintiffs' operative pleadings consist of the Second Amended Master Complaint *and* their short form complaints. ECF 117 at 3.

Mr. Zuckerberg—knew that if the broader public had access to this knowledge, other parents would make the same choice, costing Meta the young users its business depended on. This understanding is apparent in Meta's cynical strategy of marketing itself as a youth-safety oriented company—including through public statements by Mr. Zuckerberg—while aggressively discrediting any outside research showing harms to youth from its products. SAC ¶¶ 184-85, 303-304, 364. Meta's persistent focus on child safety in messaging efforts provides further evidence of the importance of such information to Plaintiffs. *See In re MyFord Touch Consumer Litig.*, 46 F. Supp. 3d 936, 957 (N.D. Cal. 2014) ("[A] reasonable jury could well conclude that the problems with the MFT system were material facts because the system arguably was the subject of Ford's marketing efforts—the system enhanced the functionality and experience of the vehicle, including its safety."); *In re Chrysler*, 295 F. Supp. 3d at 1014 ("There is, at the very least, a question of fact regarding materiality in light of the fact that the Class Vehicles were promoted as environmentally friendly in the first place. In other words, if Defendants promoted the Class Vehicles as such, they believed that such information was material to the consuming public.").

Furthermore, had Mr. Zuckerberg fulfilled his duty to Plaintiffs and made these disclosures, Plaintiffs undoubtedly would have been aware of the same. Mr. Zuckerberg's every statement and decision receives an almost unmatched degree of public attention. There are few other CEOs whose Senate testimony is posted verbatim as a story in its own right, *see* SAC ¶ 173 n.155, who can not only author an op-ed in a national paper but have their name lead the headline, *see* SAC ¶ 370 n.497, or whose every public statement is exhaustively catalogued in an academic archive, *see* SAC ¶¶179 n.175 (citing same). When Mr. Zuckerberg speaks, the whole world listens.

In addition to Mr. Zuckerberg's personal ability to command public attention, information about Meta's safety (or lack thereof) for child users generates massive public interest. *In re Chrysler-Dodge-Jeep Ecodiesel Marketing, Sales Practices, and Products Liability Litigation* provides a helpful point of comparison. In *Chrysler*, the court found that the plausibility of concealed information receiving national news coverage was enough to show that plaintiffs likely would have learned of the information had it not been concealed. 295 F. Supp. 3d at 1015. Here, such coverage is not just plausible, it actually occurred. When the information concealed by Mr. Zuckerberg was eventually revealed by whistleblower Frances Haugen, it generated national, high-profile press coverage and investigations at the highest

levels of government. SAC ¶¶ 217, 377-79. Thus, Plaintiffs would have been aware of any statements by Mr. Zuckerberg disclosing the dangers of using Meta's platforms.

For the same reasons, Plaintiffs satisfy Rule 9(b). Plaintiffs allege that Mr. Zuckerberg should have publicly revealed that Meta's platforms are harmful to children, that they would have learned of this information from resulting widespread media coverage, and that they would have behaved differently if they had known the truth about Meta's platforms. Taken together, this is more than sufficient to comply with the Rule 9(b) standard for omissions claims. *See In re Carrier IQ,Inc.*, 78 F. Supp. 3d 1051, 1113-14 (N.D. Cal. 2015) (finding that plaintiffs complied with Rule 9(b) and UCL pleading requirements because they alleged what information was omitted, that the omitted information was exclusively known by defendants, that they would have acted differently had they known the omitted information, and that the omitted information was the subject of intense public outcry); *MacDonald*, 37 F. Supp. 3d at 1096 ("Plaintiffs adequately allege the 'who what when and how,' given the inherent limitations of an omission claim. In short, the 'who' is Ford, the 'what' is its knowledge of a defect, the 'when' is prior to the sale of Class Vehicles, and the 'where' is the various channels of information through which Ford sold Class Vehicles."); *Sloan*, 287 F. Supp. 3d at 878 ("Plaintiffs would likely meet their burden under Rule 9(b) even without specifically identifying a particular advertisement in which the omitted information should have been included, so long as they provide plausible allegations . . . that they would have received the information in some way had Defendant exercised reasonable care.").

The cases cited by Mr. Zuckerberg are not to the contrary. Both *Haskins v. Symantec Corp.* and *Perkins v. LinkedIn Corp.* deal with express *misrepresentations*, not omissions. *See Haskins*, 654 F. App'x 338, 339 (9th Cir. 2016); *Perkins*, 53 F. Supp. 3d 1190, 1219-21 (N.D. Cal. 2014).[14] For an omissions claim, "[a]lthough the actual receipt of some information from a defendant might tend to demonstrate that the plaintiff had an opportunity to receive additional information, it is not necessarily the only way to establish such an opportunity." *Sloan*, 287 F. Supp. 3d at 875. Indeed, the portion of *In re ZF-TRW Airbag Control Units Products Liability Litigation*, 601 F. Supp. 3d 625 (C.D. Cal. 2022), that Mr. Zuckerberg

[14] *Tabler v. Panera LLC*, 2019 WL 5579529, at *11 (N.D. Cal. Oct. 29, 2019) and *Pirozzi v. Apple Inc.*, 913 F. Supp. 2d 840, 845 (N.D. Cal. 2012), are inapposite for the same reason.

quotes makes this distinction clear. The court held that the plaintiffs could demonstrate reliance on an omission either by showing that they had seen specific statements *or* by a more general showing—untethered from any one specific statement—that they would have been aware if the concealed information had been disclosed. *Id.* at 767.

Mr. Zuckerberg's assertion that Plaintiffs were not injured by his omissions, Mot. at 8-9, is misplaced for the same reason as his reliance argument. *See Hardee's of Maumelle, Ark., Inc. v. Hardee's Food Sys., Inc.*, 31 F.3d 573, 580 (7th Cir. 1994) (explaining that reliance "suppl[ies] the causal link between the alleged tortious act and the plaintiff's harm); *ABT Bldg. Prods. Corp. v. Nat'l Union Fire Ins. Co. of Pittsburgh*, 472 F.3d 99, 126 (4th Cir. 2006) ("[P]roximate cause may be established by evidence of reliance."); *Daniel*, 806 F.3d at 1225 ("To prove reliance on an omission, a plaintiff must show that the defendant's nondisclosure was an immediate cause of the plaintiff's injury-producing conduct."). Plaintiffs allege that use of Meta's platforms caused them a range of serious physical harms, including eating disorders, self-harm, and suicidal depression. *See e.g.*, *Booker* SFC (alleging addiction and compulsive use, anorexia, depression, anxiety, and attempted suicide). As discussed above, plaintiffs have plausibly alleged that had they known the information Mr. Zuckerberg concealed, they would have strictly limited or entirely halted their use of Meta's platforms, thereby preventing or reducing their injuries. *See supra* at 14-15.

C. Mr. Zuckerberg does not have a First Amendment right to lie to Congress.

Finally, Mr. Zuckerberg maintains that the First Amendment immunizes him from claims based upon omissions from his Congressional testimony.[15] Mot. at 11. As a threshold matter, however, Mr. Zuckerberg's First Amendment argument "constitutes an affirmative defense," which the Court may only rely upon to dismiss a complaint if it is clear from the face of the complaint that the claim is barred. *See Batis v. Dun & Bradstreet Holdings, Inc.*, 2023 WL 1870057, at *7 (N.D.Cal., 2023); *see also Acoustic Sys., Inc. v. Wenger Corp.*, 207 F.3d 287, 296 (5th Cir. 2000) ("the *Noerr-Pennington* doctrine provides only an affirmative defense against liability, not a right not to stand trial"). A plaintiff need not

[15] Mr. Zuckerberg expressly limits his *Noerr-Pennington* argument to the statements that he made in testimony to Congress. Mot. at 11. Notably, however, Plaintiffs' claims rely on many other statements by Mr. Zuckerberg that were not made in Congressional testimony. SAC ¶¶ 369(a)-(c), 370(a), (d)-(l).

plead facts to negate an affirmative defense. *Perry v. Merit Sys. Prot. Bd.*, 582 U.S. 420, 435 n.9 (2017). Here, there is no obvious bar to securing relief on the face of the complaint for at least two reasons.

First, "the knowingly false statement and the false statement made with reckless disregard of the truth, do not enjoy constitutional protection." *Garrison v. State of Louisiana*, 379 U.S. 64, 75 (1964). The *Noerr-Pennington* doctrine supplies no exception. "The doctrine derives from two Supreme Court cases holding that the First Amendment Petition Clause immunizes acts of petitioning the legislature from antitrust liability." *Theme Promotions, Inc. v. News Am Mktg. FSI*, 546 F.3d 991, 1006 (9th Cir. 2008). It does not afford greater protection than the First Amendment. *See McDonald v. Smith*, 472 U.S. 479, 485 (1985) ("there is no sound basis for granting greater constitutional protection to statements made in a petition . . . than other First Amendment expressions"). Simply stated, "neither the *Noerr-Pennington* doctrine nor the First Amendment more generally protects petitions predicated on fraud or deliberate misrepresentation[.]" *Edmondson & Gallagher v. Alban Towers Tenants Ass'n*, 48 F.3d 1260, 1267 (D.C. Cir. 1995).

To be sure, the *Noerr-Pennington* doctrine may protect deception "along the lines normally accepted in our political system." *See E.R.R. Presidents Conf. v. Noerr Motor Freight, Inc.*, 365 U.S. 127, 145 (1961); *see also Kottle v. Nw. Kidney Centers*, 146 F.3d 1056, 1061 (9th Cir 1998) ("the political arena has a higher tolerance for outright lies than the judicial arena does"). But knowingly concealing material information from Congress does not qualify. It is a crime. *See* 18 U.S.C. § 1001; *see also Allied Tube & Conduit Corp. v. Indian Head, Inc.*, 486 U.S. 492, 504 (1988) (distinguishing "misrepresentations made under oath at a legislative committee hearing" from "deceptive practices in the political arena").

Second, even if the *Noerr-Pennington* doctrine did protect false statements—as some courts have mistakenly found, *see, e.g.*, *Tuosto v. Philip Morris USA Inc.*, 2007 WL 2398507, at *5-6 (S.D.N.Y. Aug. 21, 2007)—it only applies when one "petition[s] the government for a redress of grievances." U.S. Const. Amend. I; *see Sosa v. DIRECTV, Inc.*, 437 F.3d 923, 929 (9th Cir. 2006). In the absence of evidence that a party was seeking redress from Congress, *Noerr-Pennington* does not apply. *See, e.g.*, *In re Apple Inc. Sec. Litig.*, 2020 WL 2857397, at *18 (N.D. Cal. 2020). In the present case, there is no indication in the Master Complaint or Plaintiffs' short-form complaints—whose allegations control—that Mr. Zuckerberg intended to influence any particular legislation or government effort through his

omissions and concealments. Indeed, Mr. Zuckerberg wasn't on the Hill because he wanted something from Congress, but rather because Congress wanted something from him—namely, the truth about the harm caused by Meta's platforms. *See, e.g.*, SAC ¶ 307(o) (letter from senators requesting information). Mr. Zuckerberg has been called to testify again before Congress on January 31, 2024.[16] One hopes he will finally be forthcoming about the risks Meta's platforms pose to young people. The world will be listening.

V. CONCLUSION

Plaintiffs respectfully request that the Court deny Mark Zuckerberg's motion to dismiss.

Dated: January 16, 2024

Respectfully submitted,

/s/ Previn Warren
Previn Warren
Abigail Burman
MOTLEY RICE LLC
401 9th Street NW Suite 630
Washington, DC 20004
Telephone: 202-232-5504
pwarren@motleyrice.com
aburman@motleyrice.com

Mathew P. Jasinski
Jessica C. Colombo
MOTLEY RICE LLC
20 Church Street, 17th Floor
Hartford, CT 06103
Telephone: 860-882-1681
mjasinski@motleyrice.com
jcolombo@motleyrice.com

Jonathan D. Orent
Katie Menard
MOTLEY RICE LLC
40 Westminster St., 5th Floor
Providence, RI 02903
Telephone: 401-457-7700
jorent@motleyrice.com
kmenard@motleyrice.com

Jodi Westbrook Flowers
Sara O. Couch
Jade Haileselassie
Annie Kouba
Ebony Bobbitt
Jessica L. Carroll
MOTLEY RICE LLC
28 Bridgeside Boulevard
Mt. Pleasant, SC 29464
Telephone: 843-216-9000
jflowers@motleyrice.com
scouch@motleyrice.com
jhaileselassie@motleyrice.com
akouba@motleyrice.com
ebobbitt@motleyrice.com
jcarroll@motleyrice.com

[16] U.S. Senate Committee on the Judiciary, *Durbin, Graham Announce January 2024 Hearing with Five Big Tech CEOs on their Failure to Protect Children Online* (Nov. 29, 2023), https://www. judiciary. senate.gov/press/releases/durbin-graham-announce-january-2024-hearing-with-five-big-tech-ceos-on-their-failure-to-protect-children-online.

Phyllis A. Jones (*pro hac vice*)
COVINGTON & BURLING LLP
One CityCenter
850 Tenth Street, NW
Washington, DC 20001-4956
Telephone: + 1 (202) 662-6000
Facsimile: + 1 (202) 662-6291
Email: pajones@cov.com

Attorneys for Defendants Meta Platforms, Inc., Instagram, LLC, Meta Payments, Inc., Meta Platforms Technologies, LLC, Facebook Payments, Inc., Siculus, Inc., Facebook Operations, LLC, and Mark Elliot Zuckerberg

Additional parties and counsel listed on signature pages

UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

OAKLAND DIVISION

IN RE: SOCIAL MEDIA ADOLESCENT ADDICTION/PERSONAL INJURY PRODUCTS LIABILITY LITIGATION

THIS DOCUMENT RELATES TO:

ALL ACTIONS

MDL No. 3047

Case No. 4:22-md-03047-YGR-PHK

Honorable Yvonne Gonzalez Rogers

REPLY IN SUPPORT OF MARK ZUCKERBERG'S MOTION TO DISMISS PURSUANT TO RULE 12(b)(6) THE PERSONAL INJURY PLAINTIFFS' CLAIMS

Hearing:
Date: TBD
Time: TBD
Place: Oakland, California
Judge: Hon. Yvonne Gonzalez Rogers

TABLE OF CONTENTS

Page

I. ARGUMENT 1

A. Plaintiffs Fail to Plead that Mr. Zuckerberg Owed Them a Duty to Disclose the Alleged Omissions. 1

B. Plaintiffs Fail to Establish Their Reliance on Mr. Zuckerberg's Alleged Omissions. 5

C. Plaintiffs Do Not Allege Mr. Zuckerberg Made Misleading Statements of Fact. 7

D. Mr. Zuckerberg's Congressional Testimony Cannot Give Rise to a Cause of Action. 9

II. CONCLUSION 10

TABLE OF AUTHORITIES

Page(s)

Cases

In re Apple Inc. Sec. Litig.,
2020 WL 2857397 (N.D. Cal. June 2, 2020) 8

BAC Home Loans Serv. v. Farina,
2010 Conn. Super. LEXIS 4929 (Super. Ct. June 2, 2010) 3

Bain v. Jackson,
783 F. Supp. 2d 13 (D.D.C. 2010) 3

Bank of Montreal v. Signet Bank,
193 F.3d 818 (4th Cir. 1999) 2, 5

Bays v. Hunter Sav. Ass'n,
539 F. Supp. 1020 (S.D. Ohio 1982) 4

Berger v. Sec. Pac. Info. Sys., Inc.,
795 P.2d 1380 (Colo. App. 1990) 2

Burman v. Richmond Homes Ltd.,
821 P.2d 913 (Colo. App. 1991) 5

In re Chrysler-Dodge-Jeep Ecodiesel Mktg., Sales Pracs., & Prods. Liab. Litig.,
295 F. Supp. 3d 927 (N.D. Cal. 2018) 4, 6

Coldwell Banker Whiteside Assocs. v. Ryan Equity Partners, Ltd.,
181 S.W.3d 879 (Tex. App. 2006) 3

In re Conventry Healthcare, Inc. Sec. Litig.,
2011 WL 1230998 (D. Md. Mar. 30, 2011) 8

Daniel v. Ford Motor Co.,
806 F.3d 1217 (9th Cir. 2015) 5

Dean v. Beckley,
No. CIV10-297, 2010 WL 3928650 (D. Md. Oct. 1, 2010) 3

DiMichele v. Perrella,
158 Conn. App. 726, 120 A.3d 551 (2015) 4

Eastern R.R. Presidents Conf. v. Noerr Motor Freight, Inc.,
365 U.S. 127 (1961) 9

In re Facebook PPC Advert. Litig.,
2010 WL 3341062 (N.D. Cal. Aug. 25, 2010) 6

First Choice Armor & Equip., Inc. v. Toyobo Am., Inc.,
717 F. Supp. 2d 156 (D. Mass. 2010).....2, 3

In re Fluoroquinolone Prod. Liab. Litig.,
517 F. Supp. 3d 806 (D. Minn. 2021).....2, 3

Franchise Realty Interstate Corp. v. San Francisco Loc. Joint Exec. Bd. of Culinary Workers,
542 F.2d 1076 (9th Cir. 1976).....9

Garcia v. Chrysler Grp. LLC,
127 F. Supp. 3d 212 (S.D.N.Y. 2015).....1

Haddad v. Merck & Co.,
2022 WL 17357779 (C.D. Cal. Aug. 11, 2022).....7

Jones v. Corus Bankshares, Inc.,
701 F. Supp. 2d 1014 (N.D. Ill. 2010).....7

Kaloti Enterprises, Inc. v. Kellogg Sales Co.,
2005 WI 111, 283 Wis. 2d 555, 699 N.W.2d 205.....3

Kearns v. Ford Motor Co.,
567 F.3d 1120 (9th Cir. 2009).....6

Lerner v. DMB Realty, LLC,
234 Ariz. 397, 322 P.3d 909 (Ct. App. 2014).....3

LiMandri v. Judkins,
52 Cal. App. 4th 326 (1997).....4

McCabe v. Daimler AG,
160 F. Supp. 3d 1337 (N.D. Ga. 2015).....2

McCabe v. Daimler AG,
948 F. Supp. 2d 1347 (N.D. Ga. 2013).....3

McKee v. James,
No. 09 CVS 3031, 2013 WL 3893430 (N.C. Super. July 24, 2013).....3

Metro Cable Co. v. CATV of Rockford, Inc.,
516 F.2d 220 (7th Cir. 1975).....9

Newcal Indus., Inc. v. Ikon Off. Sol.,
513 F.3d 1038 (9th Cir. 2008).....8

Nooner Holdings, Ltd. v. Abilene Vill., LLC,
668 S.W.3d 956 (Tex. App. 2023).....5

Nota Const. Corp. v. Keyes Assocs., Inc.,
45 Mass. App. Ct. 15, 694 N.E.2d 401 (1998) 2

O'Connor v. Uber Techs., Inc.,
2013 WL 6354534 (N.D. Cal. Dec. 5, 2013) 4

In re Quality Sys., Inc. Sec. Litig.,
865 F.3d 1130 (9th Cir. 2017) 8

Quashnock v. Frost,
299 Pa. Super. 9, 445 A.2d 121 (1982) 3

Roberts v. Paine,
124 Conn. 170, 199 A. 112 (Conn. 1938) 2

Sosa v. DIRECTV, Inc.,
437 F.3d 923 (9th Cir. 2006) 10

Stamm v. Salomon,
144 N.C. App. 672, 551 S.E.2d 152 (2001) 4

Stichting Pensioenfonds ABP v. Countrywide Fin. Corp.,
802 F. Supp. 2d 1125 (C.D. Cal. 2011) 7

Sunquest Info. Sys., Inc. v. Dean Witter Reynolds, Inc.,
40 F. Supp. 2d 644 (W.D. Pa. 1999) 5

In re Takata Airbag Prod. Liab. Litig.,
2017 WL 775811 (S.D. Fla. Feb. 27, 2017) 4

Tapia v. Davol, Inc.,
116 F. Supp. 3d 1149 (S.D. Cal. 2015) 1

Tuosto v. Philip Morris USA Inc.,
No. 05CIV.9384(PKL), 2007 WL 2398507 (S.D.N.Y. Aug. 21, 2007) 9

TVT Recs. v. Island Def Jam Music Grp.,
412 F.3d 82 (2d Cir. 2005) 2

White v. Lee,
227 F.3d 1214 (9th Cir. 2000) 9

In re Yahoo! Inc. Customer Data Sec. Breach Litig.,
2017 WL 3727318 (N.D. Cal. Aug. 30, 2017) 8, 9

In re ZF-TRW Airbag Control Units Prod. Liab. Litig.,
601 F. Supp. 3d 625 (C.D. Cal. 2022) 5, 6

Other Authorities

63A Am. Jur. 2d Products Liability § 779 4

I. ARGUMENT

Plaintiffs seek to impose liability on Mr. Zuckerberg based on his prominence as the founder and leader of a well-known company. Opp. 1 ("Mark Zuckerberg is not merely a CEO; he is a household name."). In effect, Plaintiffs invite the Court to rule that the normal rules governing the personal liability of corporate officers should not apply here. *See id.* at 2. The Court should reject that invitation and, applying black letter law, dismiss the claims asserted against Mr. Zuckerberg.

In their Opposition, Plaintiffs effectively concede that they do not state a claim against Mr. Zuckerberg for any ***affirmative*** misrepresentations, instead arguing that Mr. Zuckerberg is liable for alleged omissions or concealment. *See* Opp. 5 ("Plaintiffs' claims center on what Mr. Zuckerberg *failed* to say"). As explained in Mr. Zuckerberg's motion, Plaintiffs' theory fails for several reasons: (1) they do not allege Mr. Zuckerberg personally owed them a duty to disclose, (2) they do not adequately allege reliance on Mr. Zuckerberg's alleged omissions, (3) Mr. Zuckerberg's allegedly "incomplete" statements are not statements of fact upon which they can reasonably rely, and (4) Mr. Zuckerberg's statements to Congress are protected by the First Amendment. Plaintiffs' Opposition fails to meaningfully address these fatal defects.

A. Plaintiffs Fail to Plead that Mr. Zuckerberg Owed Them a Duty to Disclose the Alleged Omissions.

Plaintiffs' claim that Mr. Zuckerberg is liable for "fraudulent omissions and concealment," Opp. 7, fails as a threshold matter because Plaintiffs have not pled facts that would give rise to a duty to disclose. The Court can resolve Mr. Zuckerberg's motion on this basis alone—Plaintiffs do not, and could not, dispute that absent a duty to disclose none of their omission or concealment claims can proceed. *See, e.g.*, *Tapia v. Davol, Inc.*, 116 F. Supp. 3d 1149, 1163 (S.D. Cal. 2015) (duty to disclose an element of fraudulent concealment claim); *Garcia v. Chrysler Grp. LLC*, 127 F. Supp. 3d 212, 237 (S.D.N.Y. 2015) ("Plaintiffs have failed to allege a duty to disclose and, by extension, a plausible fraudulent concealment claim under the laws of Alabama, Florida, Georgia, New Jersey, South Dakota, and Texas.").

Plaintiffs' Opposition confirms they have not pled facts that would give rise to a legal duty to disclose between Mr. Zuckerberg personally and Meta's users. Plaintiffs point to two categories of purported "facts" allegedly giving rise to such a duty: (1) "Mr. Zuckerberg's exclusive and superior

knowledge of the ways Meta's products [allegedly] exploit minors, presenting risks to their health"; and (2) his "public, partial representations concerning the safety of Meta's products." Opp. 8. Neither supports a legal duty that would permit Plaintiffs' claims to proceed.

First, Plaintiffs' argument that Mr. Zuckerberg's "superior knowledge" created a duty to disclose ignores the bedrock principle that a duty to disclose does not exist absent a contractual or other special relationship between the parties. *See, e.g.*, *McCabe v. Daimler AG*, 160 F. Supp. 3d 1337, 1350 (N.D. Ga. 2015) (finding no duty to disclose where defendant had "no apparent relationship with Plaintiffs"); *In re Fluoroquinolone Prod. Liab. Litig.*, 517 F. Supp. 3d 806, 820 (D. Minn. 2021) ("The 'touchstone of [the Court's] duty analysis is to ask whether a plaintiff and a defendant stood in such a relationship to one another that [Illinois] law imposed upon the defendant an obligation of reasonable conduct for the benefit of the plaintiff.'").[1] "Superior knowledge" alone, absent a contractual or other special relationship, does not create a duty to disclose. If the rule were otherwise, not just corporate officers, but virtually any knowledgeable employee could be sued under a claimed legal duty to disclose to the public at large solely based on their superior knowledge of the company's operations—a remarkable expansion of liability that finds no support in the law or common sense.

Tellingly, ***none*** of the cases cited by Plaintiffs holds that a defendant had a duty to disclose absent a contractual or other special relationship with the plaintiff(s). Rather, they all involved parties in contractual or other special relationships. *See, e.g.*, *TVT Recs. v. Island Def Jam Music Grp.*, 412 F.3d 82, 91 (2d Cir. 2005) (in breach of contract suit, stating that "superior knowledge" can create a duty to disclose "In the context of a business transaction"); *Bank of Montreal v. Signet Bank*, 193 F.3d 818, 829

[1] The cases cited by Plaintiffs do not support Plaintiffs' assertion that "Mr. Zuckerberg is incorrect that 'a duty to disclose [typically] does not exist absent a contractual or other special relationship between the parties." Opp. 9. Indeed, the *Roberts* case states that "[w]hether or not there is a duty to disclose depends on the relationship of the parties," and found that there was no duty owed by the officers of a mental institution to the plaintiff patient. *Roberts v. Paine*, 124 Conn. 170, 175, 199 A. 112, 115 (Conn. 1938). The *Keyes* case only makes the generic (and correct) statement that "a duty to disclose may arise in a number of circumstances," and provides no support that a duty to disclose can exist absent a special relationship between the parties. *See Nota Const. Corp. v. Keyes Assocs., Inc.*, 45 Mass. App. Ct. 15, 19, 694 N.E.2d 401, 404 (1998). Finally, *Berger* was decided in the context of an employer/employee relationship, and similarly provides no support for the claim that a duty to disclose can exist absent a special relationship between the parties. *See Berger v. Sec. Pac. Info. Sys., Inc.*, 795 P.2d 1380, 1385 (Colo. App. 1990).

(4th Cir. 1999) ("superior knowledge" can create duty to disclose in breach of contract suit, and noting that "a duty to disclose does not normally arise when parties are engaged in an arm's length transaction"); *Coldwell Banker Whiteside Assocs. v. Ryan Equity Partners, Ltd.*, 181 S.W.3d 879, 888 (Tex. App. 2006) (stating that "the duty to disclose arises when one party knows that the other party is ignorant of the true facts and does not have an equal opportunity to discover the truth" in the context of a breach of contract suit); *Lerner v. DMB Realty, LLC*, 234 Ariz. 397, 405, 322 P.3d 909, 917 (Ct. App. 2014) (stating that "a seller may be required to disclose information when the buyer reasonably cannot discover the information for himself" in the context of a real estate sale and alleged breach of fiduciary duty by a real estate broker).[2]

Here, Plaintiffs have not pled that they had a contractual or other special relationship with Mr. Zuckerberg that could create a legal duty on his part to disclose any of the omissions alleged here. Instead, they focus solely on allegations regarding Mr. Zuckerberg's purported "superior knowledge of the harms Meta's products pose to minors." *E.g.*, Opp. 8. But Plaintiffs cite ***no authority*** that supports a personal duty for every officer to disclose under these facts, even accepting them as true. Whether a duty to disclose exists is a question that turns on the ***relationship between plaintiff and the defendant***. *See e.g.*, *In re Fluoroquinolone Prod. Liab. Litig.*, 517 F. Supp. 3d 806, 820 (D. Minn. 2021) ("The 'touchstone of [the Court's] duty analysis is to ask whether a plaintiff and a defendant stood in such a relationship to one another that [Illinois] law imposed upon the defendant an obligation of reasonable conduct for the benefit

---

[2] *See also BAC Home Loans Serv. v. Farina*, 2010 Conn. Super. LEXIS 4929, at *2 (Super. Ct. June 2, 2010) (stating "a duty to disclose arises where one party's superior knowledge of essential facts renders a transaction without disclosure inherently unfair" in context of a real estate sale) (emphasis added); *Dean v. Beckley*, No. CIV10-297, 2010 WL 3928650, at *5 (D. Md. Oct. 1, 2010) (addressing duty to disclose in context of a breach of contract dispute); *McKee v. James*, No. 09 CVS 3031, 2013 WL 3893430, at *8 (N.C. Super. July 24, 2013) (addressing duty to disclose in context of breach of contract dispute, and finding no duty to disclose); *Quashnock v. Frost*, 299 Pa. Super. 9, 23, 445 A.2d 121, 128 (1982) (addressing failure to disclose in context of sale of real estate); *First Choice Armor & Equip., Inc. v. Toyobo Am., Inc.*, 717 F. Supp. 2d 156, 162 (D. Mass. 2010) (addressing failure to disclose in context of sale of component parts to manufacturer); *Bain v. Jackson*, 783 F. Supp. 2d 13, 18 (D.D.C. 2010) (addressing duty to disclose in context of breach of contract dispute, and finding no duty to disclose); *McCabe v. Daimler AG*, 948 F. Supp. 2d 1347, 1368 (N.D. Ga. 2013) (addressing failure to disclose in context of class action brought by vehicle purchasers against vehicle manufacturers); *Kaloti Enterprises, Inc. v. Kellogg Sales Co.*, 2005 WI 111, ¶ 19, 283 Wis. 2d 555, 573, 699 N.W.2d 205, 213 (addressing failure to disclose in the context of "part[ies] to a business transaction").

of the plaintiff.'"). Plaintiffs' allegations focused solely on Mr. Zuckerberg's purported knowledge, without more, cannot create such a duty.[3]

Second, Plaintiffs' circular argument that Mr. Zuckerberg's public statements created a duty to disclose, Opp. 10-11, similarly fails. Plaintiffs again cite to inapplicable authority to support their novel—and incorrect—argument that any public statement creates a duty to disclose to the public at large. For example, Plaintiffs cite language from a treatise stating that a duty to disclose may exist "when the defendant makes partial representations," Opp. 10, but omit language from the very same discussion making clear that this is so only if the speaker "knows that the other is about to ***enter into the transaction*** under a mistake as to such facts and that the other, ***because of the relationship between them***, would reasonably expect disclosure of such facts." 63A Am. Jur. 2d Products Liability § 779 (emphasis added). The cases Plaintiffs cite all similarly involve a contractual or other special relationship between the parties. *See In re Chrysler-Dodge-Jeep Ecodiesel Mktg., Sales Pracs., & Prods. Liab. Litig.*, 295 F. Supp. 3d 927, 1008-09 (N.D. Cal. 2018) (discussing partial disclosure in context of class action brought by purchasers of trucks against manufacturers and suppliers); *In re Takata Airbag Prod. Liab. Litig.*, 2017 WL 775811, at *1 (S.D. Fla. Feb. 27, 2017) (addressing "incomplete" representations in context of personal injury suits by purchasers of vehicles against vehicle manufacturer); *LiMandri v. Judkins*, 52 Cal. App. 4th 326, 336–37 (1997) (noting that in addition to a fiduciary relationship, "[e]ach of the other three circumstances in which nondisclosure may be actionable presupposes the existence of some other relationship between the plaintiff and defendant in which a duty to disclose can arise"); *DiMichele v. Perrella*, 158 Conn. App. 726, 731, 120 A.3d 551, 554 (2015) (rejecting partial disclosure argument where parties did not "share[] a special relationship").[4] Accordingly, Plaintiffs' allegations regarding Mr. Zuckerberg's public

---

[3] Plaintiffs also argue that "Mr. Zuckerberg's duty must be considered in the context of his company's," Opp. 9, inviting the Court to impute any duty Meta might have to Mr. Zuckerberg personally. Plaintiffs do not, and could not, cite any authority in support of this argument. *See, e.g.*, *O'Connor v. Uber Techs., Inc.*, 2013 WL 6354534, at *18 (N.D. Cal. Dec. 5, 2013) (liability of corporation cannot be imputed to individual defendants merely based on their status as officers of the corporation).

[4] *See also Bays v. Hunter Sav. Ass'n*, 539 F. Supp. 1020, 1025 (S.D. Ohio 1982) (addressing partial disclosure in context of consumer credit contracts); *Stamm v. Salomon*, 144 N.C. App. 672, 680, 551 S.E.2d 152, 158 (2001) (addressing partial disclosure in context of dispute between business partners); (continued…)

statements do nothing to remedy the foundational defect in their fraudulent omission/concealment claims: Plaintiffs' have not alleged any relationship ***between Mr. Zuckerberg and Plaintiffs*** that could create a legal duty to disclose.

B. Plaintiffs Fail to Establish Their Reliance on Mr. Zuckerberg's Alleged Omissions.

To state a claim for fraudulent omission, Plaintiffs are required "to plead some facts 'to establish that they would have been aware of the [omitted fact], if it were disclosed.'" *In re ZF-TRW Airbag Control Units Prod. Liab. Litig.*, 601 F. Supp. 3d 625, 767 (C.D. Cal. 2022); *see also Daniel v. Ford Motor Co.*, 806 F.3d 1217, 1226 (9th Cir. 2015) (plaintiffs must establish "that they would have been aware of a disclosure"); Mot. 7–8. Plaintiffs fail to do so here. They devote much of their Opposition to arguing that reliance can be presumed or inferred when a material fact is omitted, and that they would have acted differently had the allegedly omitted information been revealed. *See* Opp. 12–15. This misses the point—even if Plaintiffs can show (by inference or otherwise) that they would have acted differently, they still must plead facts demonstrating that they would have actually been aware of the disclosure in the first place. Plaintiffs ***admit*** as much. *Id.* at 13 (acknowledging that "plaintiffs still must plausibly allege that they would have been aware of the omitted information had it been publicly revealed"). On this necessary element, the Complaint is silent. *See* Mot. 7–8.

The Opposition betrays the Complaint's silence on this issue by offering nothing but rank speculation and conclusory statements in an attempt to argue that Plaintiffs would have been aware of the disclosures. Plaintiffs claim without support or reasoning that "Mr. Zuckerberg's every statement and decision receives an almost unmatched degree of public attention" and that "[w]hen Mr. Zuckerberg speaks, the whole world listens." Opp. 15. These claims are supported only by allegations that some press and academic sources have compiled his statements, and that Mr. Zuckerberg has published an op-

---

Nooner Holdings, Ltd. v. Abilene Vill., LLC, 668 S.W.3d 956, 966 (Tex. App. 2023) (addressing partial disclosure in context of breach of contract claim); *Sunquest Info. Sys., Inc. v. Dean Witter Reynolds, Inc.*, 40 F. Supp. 2d 644, 657 (W.D. Pa. 1999) (rejecting plaintiff's "incomplete" representation theory where parties had an integrated contract); *Burman v. Richmond Homes Ltd.*, 821 P.2d 913, 918 (Colo. App. 1991) (addressing partial disclosure in the context of written real estate contracts, and finding that defendant had no duty to disclose); *Bank of Montreal v. Signet Bank*, 193 F.3d 818, 829 (4th Cir. 1999) (addressing partial disclosure in context of contractual dispute between banks).

ed. *See id.* (citing Compl. ¶¶ 173 n. 155, 179 n. 175, 370 n. 497). To say this amounts to an "unmatched degree of public attention" is quite the stretch. But even if it is true that Mr. Zuckerberg receives a great deal of public attention, it does not matter unless Plaintiffs can show that ***they themselves*** were listening to and relying on his statements. Yet they never allege a single instance where they read or heard a statement of his, much less that they consulted any of his statements in deciding whether to use (or to allow their minor children to use) Meta's services. *Cf. In re ZF-TRW*, 601 F. Supp. 3d at 767 (collecting cases where plaintiffs established they would have been aware of the disclosure by demonstrating that they had read specific labels or spoken to specific sales personnel).[5]

Plaintiffs also claim that past whistleblower disclosures have generated "high-profile press coverage," citing to allegations of a story published in the Wall Street Journal. Opp. 15–16 (citing Compl. ¶¶ 217, 377–79). They argue that, had Mr. Zuckerberg disclosed the alleged omissions, it would have generated similar press coverage. But this still does not establish that ***Plaintiffs themselves*** actually saw this Wall Street Journal story at the time, or would have seen similar coverage even if it had occurred. Critically, no Plaintiff alleges having actually read the Wall Street Journal story.[6]

Moreover, the information that Plaintiffs allege Mr. Zuckerberg and Meta should have disclosed is essentially the same information that they allege ***was*** in fact disclosed by the Wall Street Journal in

---

[5] Plaintiffs misleadingly suggest that the *In re ZF-TRW* court held that the plaintiffs could demonstrate reliance "by a more general showing—untethered from any one specific statement"—that they would have been aware of the disclosure. Opp. 17. The court made no such holding about a "general showing," and to the extent it concluded no one specific statement was required, it also found that the plaintiffs spoke to a specific sales representative who would have disclosed the information. 601 F. Supp. 3d at 767.

[6] *In re Chrysler-Dodge-Jeep Ecodiesel Mktg., Sales Pracs., & Prod. Liab. Litig.*, 295 F. Supp. 3d 927 (N.D. Cal. 2018), is distinguishable. While the court remarked that "[h]ad Defendants made the disclosure . . . it is plausible that the media would pick up that story, and it would have made national news," it was also the case that the plaintiffs had "interacted with and received information from sales representatives" who "would have passed on [the disclosure] to consumers at the time of the contemplated purchases." *Id.* at 1015–16. Moreover, simply presuming awareness of a disclosure from national news coverage does not comport with Rule 9(b), which requires a *particularized* showing of the circumstances surrounding how the plaintiff would have become aware of the omission. *See Kearns v. Ford Motor Co.*, 567 F.3d 1120, 1127 (9th Cir. 2009) ("[N]ondisclosure is a claim for misrepresentation in a cause of action for fraud, [so] it (as any other fraud claim) must be pleaded with particularity under Rule 9(b)."); *In re Facebook PPC Advert. Litig.*, 2010 WL 3341062, at *10 (N.D. Cal. Aug. 25, 2010) ("Plaintiffs still should be able to identify with particularity at least the specific policies and representations that they reviewed.").

2021. *See* Compl. ¶ 377 ("Meta's years-long concealment of its research ***was revealed*** just weeks later, when Frances Haugen released these studies, along with a trove of other internal Meta documents, to the Wall Street Journal." (emphasis added)). Yet *nearly all* Plaintiffs here allege that they have ***continued to use Meta's services up to the present***, when clearly they now know (because they allege it in their Complaint) of the information that they allege Mr. Zuckerberg and Meta failed to disclose. *See, e.g.*, *Booker* SFC at 4; *Garceau* SFC at 4; *B.B.* SFC at 4; *Keiser* SFC at 4; *Koisol* SFC at 4; *Jansky* SFC at 4. This fact alone belies any plausible showing of actual reliance or causation.[7] *See Haddad v. Merck & Co.*, 2022 WL 17357779, at *8 (C.D. Cal. Aug. 11, 2022) ("Plaintiffs' claims [] fail because they continued to take Singulair after the omission was cured."). In short, the fact that Plaintiffs did not change their behavior when the allegedly withheld information was disclosed effectively disproves Plaintiffs' conclusory assertions that they would have changed their behavior had Mr. Zuckerberg revealed the same information earlier.

C. Plaintiffs Do Not Allege Mr. Zuckerberg Made Misleading Statements of Fact.

While Plaintiffs argue that "Mr. Zuckerberg had a duty to disclose the [alleged] dangers Meta's products present to minors" based in part on his "his public, partial representations concerning the safety of Meta's products," Opp. 6-7, they do not dispute that—with one exception[8]—all of the statements at issue concern matters of opinion, not objective fact. *See* Mot. 9–11. Instead, they argue that even such statements can be actionable if the defendant has knowledge of facts that render the statements deceptive. In support of this contention, Plaintiffs cite only securities fraud cases where the defendants made statements about the financial health of their companies that were not true in light of known, objective facts alleged about the financial status of those companies. *See Jones v. Corus Bankshares, Inc.*, 701 F.

[7] For substantially the same reason, the omission claim fails for lack of causation. *See* Mot. 8–9. Because Plaintiffs do not establish that they would have behaved any differently had the information not been omitted, the omission cannot be the but-for cause of their injuries.

[8] The one statement that is more concrete—that Meta "do[es] not allow people under the age of 13 to sign up" for its services—is undisputedly true and therefore cannot be an actionable misrepresentation. Mot. 10–11. Plaintiffs do not respond to this point, and therefore have conceded it. *See Stichting Pensioenfonds ABP v. Countrywide Fin. Corp.*, 802 F. Supp. 2d 1125, 1132 (C.D. Cal. 2011) ("[I]n most circumstances, failure to respond in an opposition brief to an argument put forward in an opening brief constitutes waiver or abandonment in regard to the uncontested issue.").

Supp. 2d 1014, 1027–28 (N.D. Ill. 2010) (statement that company's balance sheet was "fortress-like" was actionable when the company knowingly lacked adequate reserves and had stopped originating loans); *In re Conventry Healthcare, Inc. Sec. Litig.*, 2011 WL 1230998, at *12 (D. Md. Mar. 30, 2011) (statement that health care plan was "fundamentally sound" was actionable when it knew it could not process new claims and had dwindling reserves); *In re Apple Inc. Sec. Litig.*, 2020 WL 2857397, at *16 (N.D. Cal. June 2, 2020) (statement that "business in China was very strong last quarter" was actionable because company admitted that it "'saw' troubling signs coming out of China . . . that were 'particularly bad in November'").[9]

These cases all deal with a specific kind of situation where claims are made about the financial health of a company that are ***objectively*** disprovable. Financial solvency is a numeric metric that basic accounting can assess. By contrast, concepts of "safety"—especially in the context of a novel communication service where executives must balance the competing goals of fostering free expression and restricting sensitive content—are squarely matters of opinion upon which reasonable people can disagree. *See Newcal Indus., Inc. v. Ikon Off. Sol.*, 513 F.3d 1038, 1053 (9th Cir. 2008) ("a statement that is ***quantifiable*** . . . may be an actionable statement of fact while a general, ***subjective*** claim about a product is non-actionable puffery").

As explained in the opening brief (Mot. 10), *In re Yahoo! Inc. Customer Data Sec. Breach Litig.*, 2017 WL 3727318 (N.D. Cal. Aug. 30, 2017), is squarely on point. There, the court held that statements regarding how "Defendants' prioritize the safety of their systems" were not actionable ***even though Yahoo allegedly knew of the risks of a data breach***. *Id.* at *26. Even where there is a known risk, whether a company sufficiently prioritizes safety and the prevention of such risk is inherently subjective. Plaintiffs offer no response to *In re Yahoo!* other than to misleadingly suggest that the court dismissed on other

[9] In *In re Quality Sys., Inc. Sec. Litig.*, 865 F.3d 1130 (9th Cir. 2017), the defendant "did not just describe the pipeline in subjective or emotive terms. Rather, they provided a concrete description of the past and present state of the pipeline." *Id.* at 1144. Plaintiffs do not contest that Mr. Zuckerberg's statements about safety were cast in anything but subjective or emotive terms.

grounds by citing to a different part of the opinion that dealt with a different claim.[10] In fact, the court dismissed the claims based on the safety statements "with prejudice because, as a matter of law," the statements were "mere puffery on which a reasonable consumer could not rely." 2017 WL 3727318, at *26.

D. Mr. Zuckerberg's Congressional Testimony Cannot Give Rise to a Cause of Action.

Finally, it is clear that the First Amendment's *Noerr-Pennington* doctrine immunizes Mr. Zuckerberg from claims based on statements made during the course of testimony to Congress.

Under *Noerr-Pennington*, "attempts to lobby and petition a governmental body. . . are absolutely immune" from liability. *Franchise Realty Interstate Corp. v. San Francisco Loc. Joint Exec. Bd. of Culinary Workers*, 542 F.2d 1076, 1080 (9th Cir. 1976). While the doctrine "originally arose in the antitrust context, it is based on and implements the First Amendment right to petition and therefore . . . applies equally in all contexts." *White v. Lee*, 227 F.3d 1214, 1231 (9th Cir. 2000).

Plaintiffs first mistakenly argue that the doctrine does not apply to allegedly false statements made to Congress. Mr. Zuckerberg's statements were not false, and Plaintiffs have not shown to the contrary, but in any event *Noerr* itself involved a "fraudulent" publicity campaign designed to influence legislation. *Eastern R.R. Presidents Conf. v. Noerr Motor Freight, Inc.*, 365 U.S. 127, 133 (1961). And since that time, court have repeatedly reaffirmed that when "activities occur in a legislative or other non-adjudicatory governmental setting," *Noerr-Pennington* applies even to "'conduct that can be termed unethical,' ***such as deception and misrepresentation***." *Metro Cable Co. v. CATV of Rockford, Inc.*, 516 F.2d 220, 228 (7th Cir. 1975); *see Tuosto v. Philip Morris USA Inc.*, No. 05CIV.9384(PKL), 2007 WL 2398507, at *5 (S.D.N.Y. Aug. 21, 2007) ("*Noerr–Pennington* protection has been extended to all advocacy intended to influence government action, including to allegedly false statements") (collecting cases).

Here, it is especially clear that *Noerr-Pennington* protects Mr. Zuckerberg from liability, given that Plaintiffs disavow any assertion that he made specific false statements, and instead argue only that his testimony was misleading by omission. *See* Opp. 5. Imposing liability under such circumstances

[10] The part of *In re Yahoo!* to which Plaintiffs refer considered an alleged omission from the company's privacy policy, which the court dismissed because the plaintiffs there failed to allege they read the policy. 2017 WL 3727318 at *29. This *further* supports dismissal here for the reasons explained in Part B, *supra.*

would violate the "breathing space principle" underlying the First Amendment. *See Sosa v. DIRECTV, Inc.*, 437 F.3d 923, 933 (9th Cir. 2006) (recognizing that certain allegedly false statements must be afforded protection under *Noerr-Pennington* in order to fully vindicate First Amendment rights).

Plaintiffs next argue that *Noerr-Pennington* does not apply because there "is no indication in the Master Complaint or Plaintiffs' short-form complaints" that Mr. Zuckerberg's testimony was intended to "influence any particular legislation or government effort." Opp. 18. *Noerr-Pennington*, however, has been construed broadly to apply even to "public relations campaign[s]" intended to indirectly influence legislative action, *see Sosa*, 437 F.3d at 934—conduct much less directly implicating the First Amendment's protections than live Congressional testimony. In any event, the very transcripts cited in the Complaint make clear that the hearings at which Mr. Zuckerberg testified ***did*** concern potential efforts by Congress to pass legislation.[11] These transcripts, which are incorporated by reference in the Complaint, fatally undercut Plaintiffs' dubious suggestion that Mr. Zuckerberg's Congressional testimony somehow did not involve protected petitioning activity.

II. CONCLUSION

Defendant Mark Zuckerberg respectfully requests that Plaintiffs' claims against him under Counts 8 and 9 of the Second Amended Complaint be dismissed with prejudice.

[11] *See, e,.g.*, Bloomberg Government, Transcript of Mark Zuckerberg's Senate Hearing, Washington Post (Apr. 10, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing (cited at Compl. nn.155-57) ("I have a bill . . . that would just put on the books a law that said that Facebook, and any other company that gathers information about Americans, has to get their permission, their affirmative permission, before it can be reused for other purposes. Would you support that legislation. . . ?"); *id.* ("So I have a separate piece of legislation to insure that kinds who are under 16 absolutely have a privacy bill of rights, and that permission has to be received from their parents for their children before any of their information is reused for any other purpose other than that which was originally intended. Would you support a child online privacy bill of rights for kids under 16 to guarantee that that information is not reused for any other purpose without explicit permission from the parents for the kids?"); *id.* ("[W]e talk about all these proposed legislation. . . .").

Dated: January 23, 2024

Respectfully submitted,

COVINGTON & BURLING LLP

/s/ Phyllis A. Jones

Paul W. Schmidt, *pro hac vice*
pschmidt@cov.com
Phyllis A. Jones, *pro hac vice*
pajones@cov.com
Christian J. Pistilli (*pro hac vice* pending)
COVINGTON & BURLING LLP
One CityCenter
850 Tenth Street, NW
Washington, DC 20001-4956
Telephone: + 1 (202) 662-6000
Facsimile: + 1 (202) 662-6291

Emily Johnson Henn (State Bar No. 269482)
ehenn@cov.com
COVINGTON & BURLING LLP
3000 El Camino Real
5 Palo Alto Square, 10th Floor
Palo Alto, CA 94306
Telephone: + 1 (650) 632-4700
Facsimile: +1 (650) 632-4800

Attorneys for Defendants Meta Platforms, Inc., Instagram, LLC, Meta Payments, Inc., Meta Platforms Technologies, LLC, Facebook Payments, Inc., Siculus, Inc., Facebook Operations, LLC, and Mark Elliot Zuckerberg

Social Media Victims Law Center™
600 1st Avenue
Ste 102 - PMB 2383
Seattle, WA 98104
(206) 741-4862
https://socialmediavictims.org

January 28, 2024

Meta Profits from Instagram Advertisements Targeted at Underage Kids

Introduction

In 2022, Meta Platforms earned $113.6 billion in advertising, $33.2 billion of which was from Instagram.[1] Instagram, in turn, earned $4.75 billion in advertising revenue from its 16.7 million users who are under the age of 18.[2]

In a sworn Congressional testimony, the Head of Instagram, Adam Mosseri, averred that "Instagram is built for people 13 and older. If a child is under the age of 13, they are not permitted on Instagram. When we learn someone underage has created an account, we remove them."[3] However, Mosseri's sworn statement is flatly inconsistent with the advertisements Instagram sells on its platform.

Examples of Advertising Directed to Underage Kids

Hasbro is an international toy and company with annual revenues of $5.8 billion.[4] Hasbro spent $29 million in advertising in 2022 and advertised heavily on Meta's platforms Instagram and Facebook.[5]

One of Hasbro's popular product lines is Mighty Morphin Power Rangers.

---

[1] https://investor.fb.com/investor-news/press-release-details/2023/Meta-Reports-Fourth-Quarter-and-Full-Year-2022-Results/default.aspx

[2] Raffoul A, Ward ZJ, Santoso M, Kavanaugh JR, Austin SB (2023) *Social media platforms generate billions of dollars in revenue from U.S. youth: Findings from a simulated revenue model*. PLoS ONE 18(12): e0295337. https://doi.org/10.1371/journal.pone.0295337

[3] Testimony of Adam Mosseri Head of Instagram, Meta Platforms Inc. Hearing Before the United States Senate Committee on Commerce, Science, and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security (December 8, 2021)

[4] https://hasbro.gcs-web.com/news-releases/news-release-details/hasbro-reports-fourth-quarter-and-full-year-2022-financial

[5] Julian Cannon, *Hasbro taps college athletes, and Nerfball,to build clout with Gen Z* DIGIDAY (Oct. 5. 2023)

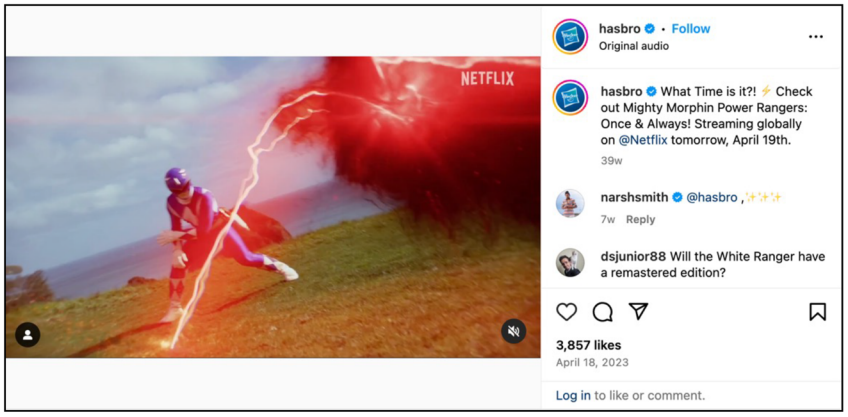


Figure 1 (https://www.instagram.com/reel/CrLiZQmMPP9/)

This advertisement for Mighty Morphin Power Rangers (*Figure 1*) is clearly aimed towards a young audience. Yet, Mighty Morphin Power Rangers received a rating of TV-Y7 from Netflix, meaning the programming is designed for ages 7 and older (*Figure 2*).



Figure 2 (https://www.netflix.com/aw/title/70184128)

Similarly, Instagram sells advertisements for Transformers: EarthSpark (*Figure 3*) and Transformers has an official Instagram account (*Figure 4*). The Transformers show is recommended for ages 8+ on Common Sense Media, with parents recommending the show to viewers of an average age of 7+ (*Figure 5*).



Figure 3 (https://www.instagram.com/reel/CaA5KB-LQL0/)

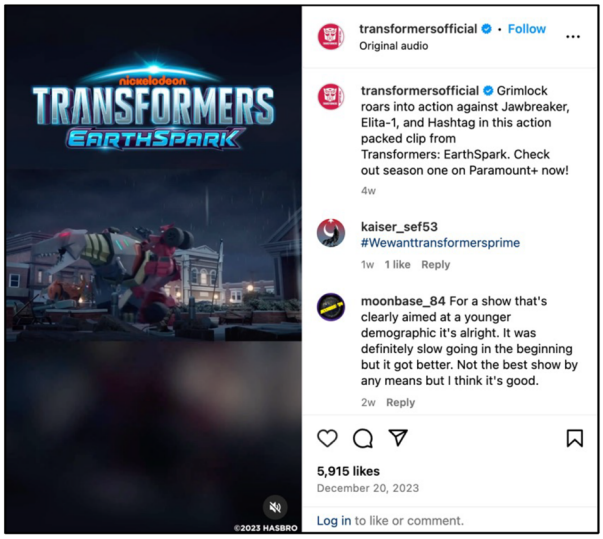


Figure 4 (https://www.instagram.com/reel/C1E_-KkrNBJ/)

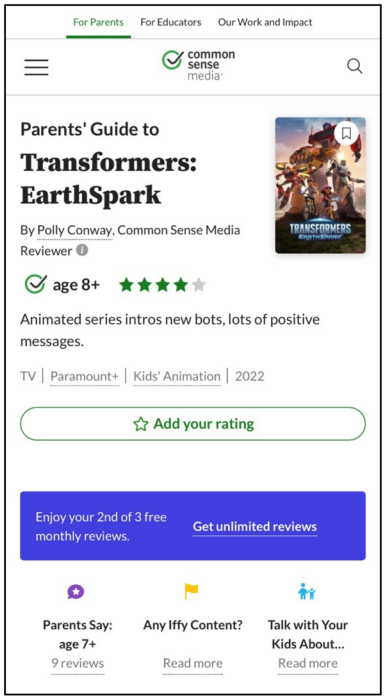


Figure 5 (https://www.commonsensemedia.org/tv-reviews/transformers-earthspark)

Another example is an Instagram advertisement promoting the Ghostbusters Ecto-1 Vehicle Playset (*Figure 6*), a toy listed for kids ages 4-years and up (*Figure 7*).

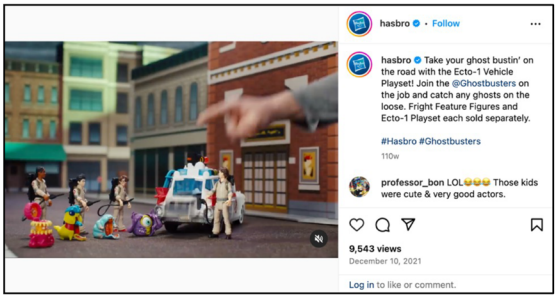


Figure 6 (https://www.instagram.com/reel/CXTl_ustnsf/)

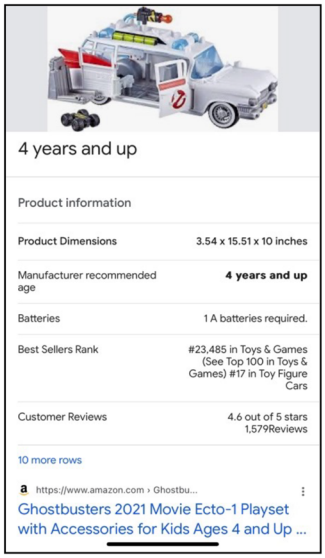


Figure 7 (https://www.amazon.com/Ghostbusters-Accessories-Collectors-Multicolor-E9563/dp/B081W1DTLR)

Hasbro also advertises its Tickle Me Elmo line on Instagram (*Figure 8*), a product listed as designed for ages 18 months to 4 years old. (*Figure 9*)



Figure 8 (https://www.instagram.com/reel/CWGbW6KjqU3/)



Figure 9 (https://www.amazon.com/Playskool-Friends-Tickle-Elmo-age/dp/B0784KMKPM)

The My Little Pony franchise developed by Hasbro provides additional examples of Instagram advertisements targeted towards children. These posts promoting the shows, My Little Pony: A New Generation (*Figure 10*) and My Little Pony: Tell Your Tale series (*Figure 11*), are both clearly catering to children. My Little Pony: A New Generation is recommended on Common Sense Media for ages 5+ (*Figure 12*).

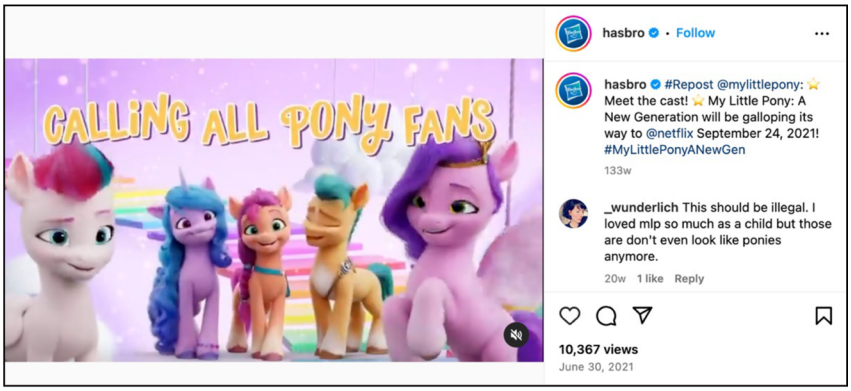


Figure 10 (https://www.instagram.com/reel/CQwaVtAjx_V/)



Figure 11 (https://www.instagram.com/reel/CqS4QURgrx6/)

Mattel is an international toy manufacturer with annual sales of $5.4 billion[6] that spent $394 million on digital advertising in 2023.[7] Like Hasbro, Mattel runs ads on Instagram targeting young children.

A post on Mattel's Instagram promotes MEGA Pokémon Action Figure Building Toys Set (*Figure 13*), a toy for ages 6-11 years old. (*Figure 14*).

[6] https://investors.mattel.com/at-a-glance/
[7] Discounted Cash Flow, *Marketing Mix Analysis of Mattel*, https://dcf.fm/blogs/blog/mat-marketing-mix#:~:text=In%202023%2C%20Mattel's%20advertising%20expenses,promote%20its%20products%20and%20brand.

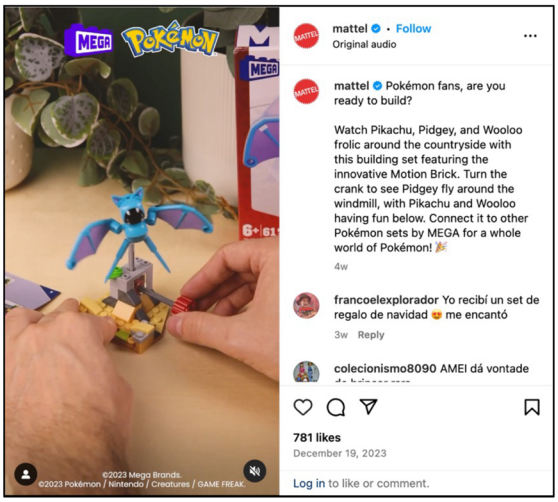


Figure 13 (https://www.instagram.com/reel/C1CwW0JJQpb/)

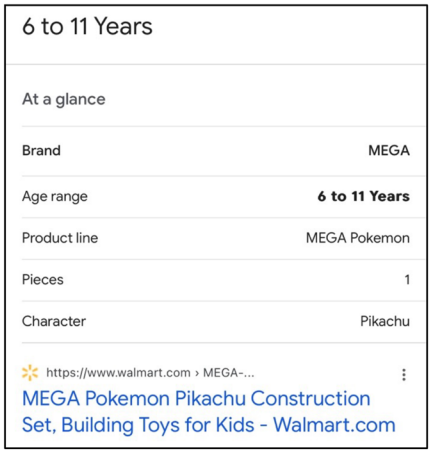


Figure 14
(https://www.google.com/search?q=mega+pokemon+construction+building+set+recommended+age)

There is also an advertisement on Instagram for Mattel's Squeeze and Blink Grogu Plushie (*Figure 15*), a toy that is recommended for ages 3 years old and up (*Figure 16*). While older fans may see the plushie as a collectable, the Instagram ad is clearly geared towards children.

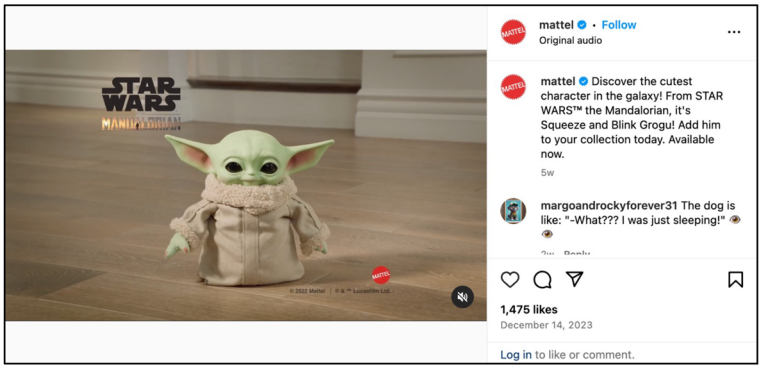


Figure 15 (https://www.instagram.com/reel/C02aAREqMu7/)

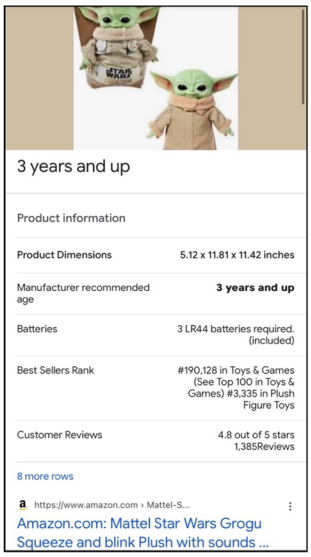


Figure 16 (https://www.amazon.com/Mattel-Squeeze-sounds-movement-Collectible/dp/B09NZD9S1M)

Mattel also advertises on Instagram for the Cuutopia Star Wars Plushie (*Figure 17*), a toy that is recommended for ages 3 years old and up (*Figure 18*).

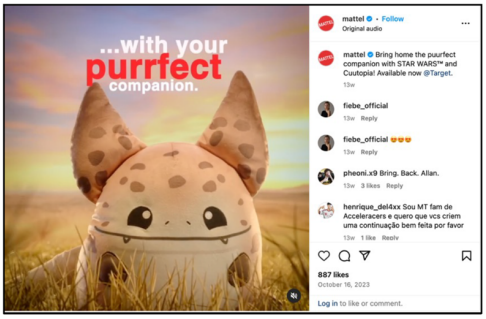


Figure 17 (https://www.instagram.com/reel/CyeI8jFOwie/)

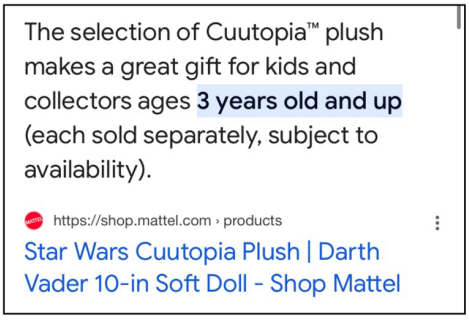


Figure 18 (https://shop.mattel.com/products/star-wars-cuutopia-10-inch-darth-vader-plush-hll03#:~:text=The%20selection%20of%20Cuutopia%E2%84%A2,separately%2C%20subject%20to%20availability).)

Conclusion

Although Meta tells the public that Instagram does not target underage kids, its conduct proves otherwise. Meta earns billions of dollars selling advertisements on Instagram that are unquestionably designed to kids under the age of 13. While Meta may not explicitly encourage these ads, it has clearly chosen to turn a blind eye to underage use of Instagram rather than turn down the revenue from toy manufacturers whose advertisements target these users.