Meta

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.

<u>Questions from Senator Whitehouse</u>

***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.