1 Hacker Way
Menlo Park, CA 94025
United States

March 21, 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:

Thank you for the questions for the record from the Senate Judiciary Hearing entitled “Big Tech and the Online Child Sexual Exploitation Crisis” on January 31, 2024. As discussed with Committee staff, due to the voluminous number of requests, Meta will provide responses to the questions on a rolling basis. Enclosed are Meta’s initial responses.

Sincerely,

Meta Platforms, Inc.

Questions from Senator Durbin

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,[1] Instagram,[2] and Messenger.[3]

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 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,

[1] How do I report an abusive photo on Facebook? | Facebook Help Center
[2] How to Report Things | Instagram Help Center
[3] Reporting Conversations | Messenger Help Center (facebook.com)

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.[4] 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. These notices help people avoid scams, spot impersonations and, most importantly, flag potentially suspicious accounts attempting to connect to minors. In just one month in 2023, more than 85 million people saw safety notices on Messenger.

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.

[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.

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.

Questions from Senator Graham

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,[5] Instagram,[6] and Messenger.[7]

Finally, 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

[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)

teen accounts to these accounts, or vice versa.[8] 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 just one month in 2023, more than 85 million people saw safety notices on Messenger.

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.

[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.

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.

Questions from Senator Cruz

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 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.

Questions from Senator Hawley

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.

Questions from Senator Lee

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,[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, and we do not recommend

[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)

teen accounts to these accounts, or vice versa.[12] 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 just one month in 2023, more than 85 million people saw safety notices on Messenger.

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.

[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.

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,[13] Instagram,[14] and Messenger.[15]

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. In just one month in 2023, more than 85 million people saw safety notices on Messenger.

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.

[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)

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 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.

Questions from Senator Padilla

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,[16] Instagram,[17] and Messenger.[18]

Finally, 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.[19] 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 just one month in 2023, more than 85 million people saw safety notices on Messenger.

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

[16] How do I report an abusive photo on Facebook? | Facebook Help Center
[17] How to Report Things | Instagram Help Center
[18] Reporting Conversations | Messenger Help Center (facebook.com)
[19] Meta identifies adult accounts "exhibiting potentially suspicious" behavior using numerous signals, including for example, having been recently blocked or reported by a young person.

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.

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 30 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 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.