**Senator Dick Durbin**
**Chair, Senate Judiciary Committee**
**Written Questions for Linda Yaccarino**
**Chief Executive Officer, X Corp.**
**February 7, 2024**

1. For each year from 2019 to 2023, please provide the following:

   a. the total number of users on your platform;

      **2019 - Average DAU (Daily Active Users) was approximately 152 million**
      **2020 - Average DAU was approximately 192 million**
      **2021 - Average DAU was approximately 217 million**
      **2022 - Average DAU was approximately 238 million**
      **2023 - Average DAU was approximately 245 million**

   b. the total number of users under the age of 18 on your platform;

      **2019 - 2022 - We will follow up with this figure.**
      **2023 - As of January 2024, Average DAU of users 13-17 in US is approximately 650,000**

   c. the estimated number of users under the age of 13 on your platform;

      **X does not allow individuals under the age of 13 to open an account on the platform. We do not have estimates for the number of users under the age of 13.**

   d. your company's annual revenue;

      **Revenue for the years 2019, 2020, 2021 and first two quarters of 2022 were disclosed to the SEC as part of Twitter's public company filings.**

      **2019 revenue was $3.46 billion, with an operating income of $366 million, and net income of $1.47 billion.**

      **2020 revenue was $3.72 billion, with an operating income of $27 million, and net loss of $1.12 billion.**

      **2021 revenue was $5.08 billion, with a 2021 operating loss of $493 million and net loss of $221 million.**

      **2022 Q1 revenue was $1.2 billion, an operating loss of $128 million, and a net income of $513 million which includes a pre-tax gain of $970 million from the sale of MoPub for $1.05 billion and income taxes related to the gain of $331 million.**

      **2022 Q2 revenue was $1.18 billion, with an operating loss of $344 million, and a net**

**loss of $270 million.**

**Beginning in Q3 2022, X Corp., a private company, took over operation of the platform. As a privately-held company, X does not maintain or release public financial statements.**

e. your company's annual budget for trust and safety;

**X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.**

f. your company's annual budget to address online child sexual exploitation;

**X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.**

g. the total number of individuals at your company working to address trust and safety broken down between employees and contractors;

**We have approximately 2,300 people who work on Trust & Safety and content moderation.**

h. the total number of individuals at your company working to address online child sexual exploitation broken down between employees and contractors.

**X has a combination of program managers, policy specialists, operations specialists, engineers, legal professionals, government affairs professionals, and other functions that work on issues related to child safety.**

2. How did your company determine that 13 was the appropriate age for a child to begin using your platform?

**Our age limit is in alignment with the Children's Online Privacy and Protection Act.**

3. What legal obligation does your company have in the United States to ensure that your platform or features of your platform are safe for children before they are launched?

**The Children's Online Privacy and Protection Act, as well as guidance from the Federal Trade Commission, provide a framework for ensuring safety and privacy of minors online.**

4. For users under the age of 18,

**a.** what are the default privacy settings for their accounts?

**Accounts belonging to known minors will be defaulted to a "protected" setting. This**

**means that known minors will receive a request when new people want to follow them (which they can approve or deny), that their posts will only be visible to their followers, and that their posts will only be searchable by them and their followers (i.e. they will not appear in public searches). Under this setting, accounts belonging to known minors will be restricted to receiving DMs from accounts they follow by default. We also utilize an age lock. Once a new user enters a date of birth that makes them under the age of 18, they will be stopped from re-entering a new date of birth for that account.**

**We also take steps to limit exposure to sensitive content. Known minors or viewers who do not include a birth date on their profile are restricted from viewing specific forms of sensitive media such as adult content. X obscures sensitive media behind notices and interstitials. This includes our product age restrictions that restricts known minors from viewing adult content.**

**In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.**

**More information on our protected account settings can be found in our Help Center: https://help.twitter.com/en/safety-and-security/public-and-protected-posts.**

**b.** what limitations are placed by default on content these users can access, content that will be directed toward them, and individuals they can communicate with?

**See answer above.**

**c.** can they change their default settings without the awareness of their parent or guardian, or without the consent of their parent or guardian?

**Yes.**

**d.** In 2023, how many changed their default settings?

**In response to your inquiry, our teams are investigating this question and will be happy to follow up with you.**

**5.** If default settings are different for users aged 16 and 17 than they are for users under the age of 16, please explain why these groups of users are treated differently, how the decision to treat these groups of users differently was made, and whether any company personnel voiced objections to or raised concerns about the differing treatment of these groups of users.

**The default settings are not different for users between the ages of 13-17.**

6. Please describe what parental controls, if any, are available on your platform. What studies, research, summaries, or data does your company have reflecting the efficacy of its parental controls and child safety measures? Please provide these studies, research, summaries, or data.

   **X does not currently have parental controls, however, we will be engaging with parents and parents groups to solicit feedback on what tools could be helpful to develop. While there are not a great amount of kids and young teens on X, there are a lot of parents that we can learn from and involve in designing new products and solutions.**

7. **Concerning international law,**

   a. what steps have your company and its subsidiaries taken to comply with the European Union's *Digital Services Act*?

      **A number of steps have been taken with regard to the Digital Services Act, including risk assessments, compliance process, introduction of a specific reporting process, the publication of an EU transparency report, and processes relating to academic data access. We continue to work closely with the European Commission to detail our compliance with the DSA.**

   b. what steps have your company and its subsidiaries taken to comply with the United Kingdom's *Online Safety Act*?

      **The Online Safety Act has not yet come into force, but we are engaging closely with the regulator, OFCOM, and are evaluating any potential policy and product changes required.**

   c. what steps have your company and its subsidiaries taken to comply with Australia's *Online Safety Act*?

      **X complies with the Online Safety Act and continues to work diligently to cooperate in good faith with the Australian eSafety Commissioner. Given ongoing litigation, we decline to comment further at this time.**

   d. if those laws create a safer, healthier online experience for kids on your platforms, do you commit to implement these changes for users in the United States? If not, why not?

      **In many areas, the work that we do to comply with these laws will have a global impact, for example our investments in tackling child sexual exploitation content. Improvements in our reporting flow and greater transparency relating to our content moderation work have been rolled out globally.**

8. X claims to prioritize reducing child exploitation on its platform. In fact, in November 2022, Elon Must tweeted "Priority #1" in response to a tweet about the company addressing child sexual exploitation content.

Yet, within the past two years, X reduced its global trust and safety staff by 30 percent, including 80 percent of its staff engineers. The head of X's trust and safety team also resigned from the company in 2023.

While this was happening, *NBC News* reported that "at least dozens of accounts have continued to post hundreds of tweets in aggregate using terms, abbreviations and hashtags indicating the sale of... child sexual exploitation material." Researchers at Stanford University similarly found that X failed to prevent dozens of known images of child sexual abuse from being posted on its platform in recent months.

How is X prepared to fight child exploitation on its platform with an understaffed and under-resourced trust and safety team?

**Since acquisition, we have invested in technology, training, and people to strengthen our approach to combating child sexual exploitation on X. In 2023, as a result of our investment in additional tools and technology to combat CSE, X suspended 12.4 million accounts for violating our CSE policies. This is up from 2.3 million accounts in 2022.**

**Not only are we detecting more bad actors faster, we are also building new defenses that proactively reduce the discoverability of posts that contain this type of content. One such measure that we have recently implemented has reduced the number of successful searches for known Child Sexual Abuse Material (CSAM) patterns by over 99% since December 2022.**

**We are investing in products and people to bolster our ability to detect and action more content and accounts, and are actively evaluating advanced technologies from third-party developers that can enhance our capabilities.**

**In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year. For the first time ever, we are evaluating all videos and GIFs posted on X for CSAM. Since launching this new approach in July 2023, we have matched over 70,000 pieces of media.**

**We are more vigilant and aggressive than ever in our enforcement. Our team regularly reviews and implements improvements to the measures we take to combat online child sexual exploitation to ensure their ongoing efficacy and performance. Our increased investment in this area throughout the year has yielded significant, measurable results.**

**Since April, we have increased training for content moderators on the tools and policies for NCMEC reporting. In turn, this has led to a 10x increase in the volume of manually-submitted NCMEC reports, from an average of 6,300 reports per month to an**

**average of 64,000 reports per month from June through November 2023. We are evaluating more sources of potential CSAM than we could before.**

**While reports focus on the cuts we made over a year ago, we are hiring across the company and we are excited about building a Trust & Safety Center of Excellence in Austin, Texas. We have a goal of hiring 100 Trust & Safety agents, moving more capacity in-house, relying less on contractors. This shift will increase efficacy, decrease turnover, develop more specialization across issue areas, improve quality assurance, and strengthen our enforcement. We want to work with Congress and all stakeholders to help develop our capacity here in the United States for this important work–we stand ready to partner on this workforce development challenge.**

9. At the hearing, you repeatedly said that X is not a platform of choice for youth. However, you recently stated at a forum that Gen Z is X's fastest-growing demographic, with 200 million teenagers and young adults visiting the platform each month. You also recently noted that 70 percent of X's growth in the prior six months was driven by Gen Z users.

   What is your company doing to prevent the grooming of young people joining X?

   **The growth of Gen Z users is largely in the older portion of the generation in their 20s, as our user base of minors between the ages of 13-17 has decreased over the last year.**

   **We are currently beta testing a new text-based machine-learning classifier to detect different types of child abuse discussion, which has yielded enforcement for grooming behavior. We look forward to sharing more as we continue to develop and implement this technology.**

10. In an October meeting, Elon Musk highlighted a new feature on X that allows paying users to upload hours-long videos. You recently said that "the strides we are making so quickly in video" are "certainly getting everyone's attention."

    What steps has X taken to ensure its new video features are safe for the growing number of young people on your platform?

    **People use X to show what is happening in the world, often sharing images and videos as part of the conversation. Sometimes, this media can depict sensitive topics, including graphic content, adult nudity, and sexual behavior. We recognize that some people may not want to be exposed to sensitive content, which is why we balance allowing people to share this type of media with helping people who want to avoid it to do so.**

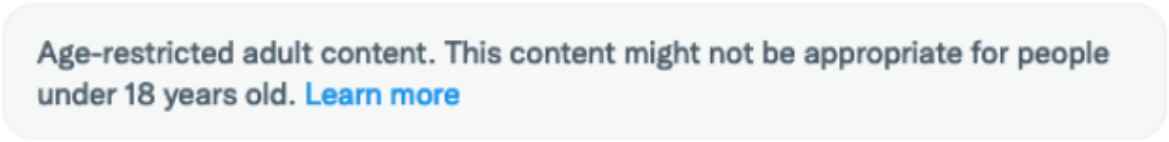
    **For this reason, you can't include graphic content, adult nudity, or sexual behavior within areas that are highly visible on X, including in live video, your profile picture or header, List banners, or Community cover photos. If you share this content on X, you need to mark your media or your account as sensitive. Doing so places images and videos behind a content warning that needs to be acknowledged before your media can be viewed. Using this feature means that people who don't want to see sensitive media**

**can avoid it, or make an informed decision before they choose to view it. We also restrict graphic media, adult nudity, and sexual behavior for viewers who are under 18 or viewers who do not include a birth date on their profile. Beginning January 2024, you may begin to see new media content warnings on posts that X has designated as Graphic (containing violent or hateful imagery) or containing Adult media (adult nudity and sexual behavior). When these new content warnings are available for you to use, please be sure to continue marking your sensitive media accordingly.**

**Under this policy, there are also some types of sensitive media that we don't allow at all, because they have the potential to normalize violence and cause distress to those who view them.**

**We restrict viewers who are under 18, or who do not include a birth date on their profile, from viewing adult content.**

**Below is an example of the interstitial we use to restrict sensitive content:**




11. An October 2023 report from Australia's eSafety Commissioner revealed that X does not employ language analysis technology to detect likely online grooming. This sets the company apart from other tech companies like Google, TikTok, and Twitch. In defense, X claimed that, although it continues to monitor the development of such technology, it is not of "sufficient capability or accuracy to be deployed by [X]."

    Could you elaborate on why language analysis technology is not capable or accurate for use on X when other platforms have demonstrated otherwise?

    **X is currently beta testing a text-based machine learning classifier developed by Thorn to assist in the detection of sextortion, CSAM, child-access, and child sexual abuse discussion.**

    **Last year, following our submission to the Australian regulator, we developed and implemented our first CSE text model. The CSE model detects text-based posts that discuss CSE. We launched this model in February 2023 and it has been critical in addressing CSE spam. Since launching, we have restricted its impact to only a subset of languages and users due to several false positives, but it continues to be our most flexible means of CSE text detection with 65,000 total suspensions so far.**

12. An impacted parent provided a statement to the Committee. She describes reporting CSAM depicting her son to X, only to be told that the images did not violate their policy. When she sued X, the lawsuit was dismissed because of Section 230.

When an individual reports child sexual abuse material on X, how do you resolve disputes as to whether the content violates X's policies? How long does that process take?

**Anyone can report potential CSAM, whether they have an X account or not.**

**In the majority of cases, the consequence for violating our CSAM policy is immediate and permanent suspension from the platform. In addition, violators will be prohibited from creating any new accounts in the future.**

**Additionally, when we are made aware of content depicting or promoting child sexual exploitation, including links to third party sites where this content can be accessed, we immediately remove it without further notice and report to the National Center for Missing & Exploited Children (NCMEC) where appropriate.**

**In a limited number of situations, where we haven't identified any malicious or sexually exploitative intent, we will require the user to remove this content. We will also temporarily lock the user out of their account before being able to post again. Further similar violations lead to the account being permanently suspended.**

**We review 100% of reports for child sexual exploitation and take immediate action on confirmed hash matches. For media-based violations, since it is visual in nature, it is usually easy to identify without needing language expertise. For text-based violations we have dedicated employees with different language expertise who are reviewing both written and media-based content. We have dedicated training resources for agents who review CSAM.**

**We have quality assurance processes and dedicated specialists that help us identify gaps in policies and enforcement. Our content moderators provide moderation services 24 hours a day, 7 days a week. We have teams spread around the world specifically trained in this highly sensitive and complex topic so that we can provide the best possible level of coverage in the languages we serve on X. For safety and security reasons, the locations of these teams are not disclosed. We have a dedicated tool that not only takes action on content, but also has the ability to dispatch communication to NCMEC. This allows us to report CSAM to NCMEC as fast as possible.**

13. You testified that X has a zero-tolerance policy for child sexual exploitation and that users who violate this policy face immediate and permanent suspension. Last summer, X suspended the account of an individual who tweeted an image of a toddler being tortured. When the account was suspended, Elon Must tweeted that the account, "was suspended for posting child exploitation pictures." Four days later, X reinstated the account. The individual who produced that child sexual abuse material was later sentenced to 129 years in prison for sexually abusing children as young as 18 months old. The image has now drawn more than 3 million views and 8,000 retweets.

    Please explain how this reflects X's zero-tolerance policy.

**Our public policy sets out how in the rare circumstances where content is shared by an account from the perspective of outrage or to raise awareness, we will remove the content and give the account holder a final warning. Any further violations will result in permanent suspension.**

**We followed our publicly stated policy, and the user was given a final warning. The user was also reported to NCMEC, which triggered the initial suspension.**

**Senator Lindsey O. Graham**
**Questions for the Record**
**Ms. Linda Yaccarino, CEO, X Corp.**
**"Big Tech and the Online Child Sexual Exploitation Crisis"**
**January 31, 2024**

1. Do you support S. 1207, the bipartisan EARN IT Act? Why or why not?

**We support the concepts within EARN IT that encourage the development of best practices to combat the distribution of CSAM. We welcome more mechanisms for industry, law enforcement, and government to share these best practices and increase collaboration. We do not believe that Section 230 protections should be conditioned on the implementation of these best practices. We also have concerns about creating an unelected commission that can become politically charged with the changing of administrations.**

2. What measures are you taking to prevent and address sextortion, including financial sextortion, on your companies' platforms?
   a. What methods are in place to detect and disrupt this type of abuse in real time?

      **We use a mixture of proactive detection, user reporting, and human capacity to enforce our private information and media policy.**

      **You may not publish or post other people's private information without their express authorization and permission. We also prohibit threatening to expose private information or incentivizing others to do so.**

      **In addition, you may not share private media, such as images or videos of private individuals, without their consent. However, we recognise that there are instances where users may share images or videos of private individuals, who are not public figures, as part of a newsworthy event or to further public discourse on issues or events of public interest. In such cases, we may allow the media to remain on the platform.**

      **Sharing someone's private information online without their permission, sometimes called doxxing, is a breach of their privacy and of the X Rules. Sharing private information can pose serious safety and security risks for those affected and can lead to physical, emotional, and financial hardship.**

      **When reviewing reports under this policy, we consider a number of things, including:**

      - ***What type of information is being shared***
        - **We take this into consideration because certain types of private or live information carry higher risks than others, if they're shared without permission. Our primary aim is to protect**

**individuals from potential physical harm as a result of their information being shared, so we consider information such as physical location and phone numbers to be a higher risk than other types of information. We define "live" as real-time and/or same-day information where there is potential that the individual could still be at the named location.**

- ***Who is sharing the information***
    - **We also consider who is sharing the reported information and whether or not they have the consent of the person it belongs to. We do this because we know that there are times when people may want some forms of their personal information to be shared publicly. For example, sharing a personal phone number or email for professional networking or to coordinate social events or publicly sharing someone's home addresses or live locations to seek help after a natural disaster.**
- ***Whether the information available elsewhere online***
    - **If the reported information was shared somewhere else before it was shared on X, e.g., someone sharing their personal phone number on their own publicly accessible website, we may not treat this information as private, as the owner has made it publicly available. Note: we may take action against home addresses being shared, even if they are publicly available, due to the potential for physical harm.**
- ***Why the information is being shared***
    - **We also factor in the intent of the person sharing the information. For example, if we believe that someone is sharing information with an abusive intent, or to harass or encourage others to harass another person, we will take action. On the other hand, if someone is sharing information in an effort to help someone involved in a crisis situation like in the aftermath of a violent event, we may not take action. Note: regardless of intent, if the information is not shared during a crisis situation to assist with humanitarian efforts or in relation to public engagement events, we will remove any posts or accounts that share someone's live location.**
- ***Sharing private media***
    - **Posting images is an important part of our users' experience on X. Where individuals have a reasonable expectation of privacy in an individual piece of media, we believe they should be able to determine whether or not it is shared. Sharing such media could potentially violate users' privacy and may lead to emotional or physical harm. When we are notified by individuals depicted, or their authorized representative, that they did not consent to having media shared, we will remove the media. This policy is not applicable to public figures.**

**What is in violation of this policy?**

**Under this policy, you can't share the following types of private information, without the permission of the person who it belongs to:**

- **home address or physical location information, including street addresses, GPS coordinates or other identifying information related to locations that are considered private;**
- **live location information, including information shared on X directly or links to 3rd-party URL(s) of travel routes, actual physical location, or other identifying information that would reveal a person's location, regardless if this information is publicly available;**
- **identity documents, including government-issued IDs and social security or other national identity numbers – note: we may make limited exceptions in regions where this information is not considered to be private;**
- **contact information, including non-public personal phone numbers or email addresses;**
- **financial account information, including bank account and credit card details;**
- **other private information, including biometric data or medical records;**
- **media of private individuals without the permission of the person(s) depicted; and**
- **media depicting prisoners of war posted by government or state-affiliated media accounts on or after April 5, 2022.**

**The following behaviors are also not permitted:**

- **threatening to publicly expose someone's private information;**
- **sharing information that would enable individuals to hack or gain access to someone's private information without their consent,e.g., sharing sign-in credentials for online banking services;**
- **asking for or offering a bounty or financial reward in exchange for posting someone's private information;**
- **asking for a bounty or financial reward in exchange for not posting someone's private information, sometimes referred to as blackmail.**

3. Please provide the committee statistics on how long it takes your company to respond to various types of legal process from law enforcement?

**X endeavors to respond to legal process received from law enforcement and appropriate government entities in a prompt manner. Specifically, X endeavors to respond prior to the enumerated production date required by law in the particular jurisdiction or outlined in the individual legal process.**

4. Do you notify your users when law enforcement serves subpoenas/summons for subscriber information and specifically requests not to notify the subscriber/user?

   **Upon receipt of a law enforcement request that includes a valid non-disclosure order, X does not notify the user unless permitted by law or court order to do so.**

5.
   a. If you notify the subscriber, how long do you wait until notification goes out?

      **For purposes of transparency and due process, X's policy is generally to notify users of requests for their X account information prior to disclosure of said account information.**

   b. Are you aware that by notifying the subscriber about a law enforcement subpoena for their subscriber information that you are jeopardizing critical evidence that could be erased before law enforcement can serve warrants?

      **Prior to notifying a user regarding a law enforcement request, X ensures that the requested data is preserved. Doing so, mitigates any risk of data loss.**

   c. Would your company agree to a 90-day non-disclosure to subscribers to allow law enforcement ample time to secure proper legal process?

      **To request data from X, law enforcement must seek and obtain proper legal process, thereby offering an opportunity to seek a non-disclosure order.**

Do you actively seek out and incorporate feedback and insight from survivors of online sexual exploitation to improve your trust and safety policies and practices and to prevent and disrupt child sexual abuse material (CSAM) production and distribution on your platform? Can you provide examples?

**We have received feedback and input from representatives of the End Online Sexual Exploitation and Abuse of Children Coalition. We welcome introductions to any other survivors or survivors groups that your office has connections to.**

   d. If not, please explain.

6. During our hearing, you testified that you collaborate with parents and parent organizations to create mechanisms to keep children safe online. Please elaborate and cite examples of your company's work with non-employee parents and parent organizations.

**Historically, in the US, we collaborated with Parents Together on product feedback, policy, enforcement, and public policy. We commit to developing a broader set of relationships with parents and parent organizations.**

7. Why does your company have the age limit of 13 years old for a user to sign up for an account?

**Our age limit is in alignment with the Children's Online Privacy and Protection Act.**

   a. Why not younger or older?

      **To the best of my knowledge, we have not considered allowing users under 13 open accounts. In some countries internationally, the law has been set higher and we have adhered to those laws.**

8. How many minors use your platform? How much money does your company make annually from these minors?

**As of December 2023, in the United States, there are approximately 650,000 daily active users between 13-17.**

9. What percentage of your employees work on trust and safety and how much money does your company invest annually in trust and safety?

**X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.**

10. It is sometimes challenging for law enforcement conducting criminal investigations to determine the true identity of a person behind a name on social media or other online platforms, and whether an online identity is an actual person. What are you doing to validate the true identity of users – or the fact that a user is a human – when they create an account on your platforms?

**X uses a variety of proprietary methods to analyze account signals in order to determine the authenticity of a user. In addition, X utilizes Premium subscriptions as a method to authenticate humans because we are able to collect payment information and identification.**

**X uses ID verification for:**

**(1) User Experience Enhancement: X will provide a voluntary ID verification option for certain X features to increase the overall integrity and trust on our platform. We collect this data when X Premium subscribers optionally choose to apply for a verified badge by verifying their identity using a government-issued ID. Once confirmed, a verified label is added to the user's profile for transparency and potentially unlocking additional benefits associated with specific X features in the future. This option is currently only available to individual users and not businesses or organizations.**

**(2) Safety and Security Purposes: In certain instances, X may require your government-issued ID when needed to ensure the safety and security of accounts on our platform. We collect this data when investigating and enforcing our policies and may**

**request an ID verification in response to impersonation reports. Currently, X focuses on account authentication to prevent impersonation, and may explore additional measures, such as ensuring users have access to age-appropriate content and protecting against spam and malicious accounts, to maintain the integrity of the platform and safeguard healthy conversations.**

11. Is your company using safety technology to detect and prevent live video child sexual abuse on your platforms and apps that allow users to stream or share live video? If not, please explain.

**Yes, we utilize safety technology to detect and prevent child abuse in our livestream product.**

    a. Has your company tested that or similar technology? If not, are you developing similar technology to address child sexual abuse in live video?

    **We utilize technology that detects nudity and presence of children in order to detect possible CSAM livestreams. X does not allow adult content in live video.**

12. How are you measuring if your trust and safety policies, practices, and tools are effective in protecting children from sexual abuse and exploitation on your platform?

    **We are constantly evaluating the effectiveness of our policies, tools, and enforcement on the safety of our users and the integrity of our service. We want to make X the most inhospitable place for bad actors seeking to exploit children. We evaluate the accuracy of our detection, the speed of our action, the quality of our reporting to NCMEC, the integrity of our data, the efficacy of automated interventions, and the productivity of our partnerships.**

    a. What specific metrics or key performance indicators do you use?

    **We evaluate a range of indicators, such as action rates, reporting rates, automated v. manual reporting, response rates to user reports, detection rates, to name a few.**

13. Is your company using language analysis tools to detect grooming activities? If not, please explain.

a. What investments will your company make to develop new or improve existing tools?

    **We are currently beta testing a new text-based machine-learning classifier to detect different types of child abuse discussion, which has yielded enforcement for grooming behavior. We look forward to sharing more as we continue to develop and implement this technology.**

14. What resources have you developed for victims and survivors of abuse on your platforms?

**The uniqueness of X is the role it serves as a platform for public conversation, the global town square of the internet. X has always been a place for victims to bring awareness to their causes and issues of public concern, like legislation. We will continue to support organizations around the world that promote online safety and we welcome any recommendations of victims groups that we could support in their campaigns and advocacy.**

15. What is your response to requests for content removal from CSAM survivors and other members of the public?

    **We review reports of CSE, private information, and non-consensual nudity from users and the public, as our reporting forms are available to anyone, whether or not you have an account.**

16. While you mentioned several times throughout the hearing that X Corp. is a new company, it is no secret that X is Twitter rebranded. At its peak, how many trust and safety employees did Twitter have on staff and how many trust and safety employees are on staff today at X?

    **We have approximately 2,300 people who work on Trust & Safety and content moderation.**

17. What resources does X dedicate to its child safety team and how has this team been stabilized following X's larger corporate changes over the past two years?

    **We have a mix of agents, policy leads, program managers, engineers, legal professionals, and government affairs professionals who work on issues of child safety, which has remained consistent over the past two years.**

18. Why does X not participate in NCMEC's "Take It Down" program to help stop the sharing of and remove nude and sexually explicit photos of minors?

    **Our teams had recently prioritized the Tech Coalition's Project Lantern and are now evaluating the technical requirements of the program.**

19. What voluntary hash-sharing or other information sharing initiatives does X participate in to help combat child sexual exploitation?

    **We participate in hash-sharing via NCMEC and the Technology Coalition. We have applied to the Tech Coalition's Project Lantern information sharing program. We are also evaluating participation in StopNCII.org and the NCMEC Take It Down program. We are also members of the Internet Watch Foundation and work with Thorn, in addition to international NGOs.**

**Senate Judiciary Committee Hearing**
**"Big Tech and the Online Child Sexual Exploitation Crisis"**
**Questions for the Record**
**for Linda Yaccarino**
**Submitted February 7, 2024**

## QUESTIONS FROM SENATOR SHELDON WHITEHOUSE

1. What exemptions from the protections of Section 230 would your company be willing to accept?

**Section 230 is often referred to as the "26 words that made the internet" with good reason. The free, open internet would not exist in a world in which users and websites face liability for merely disseminating the speech of others. Without it, free speech on the internet would cease to exist, as websites would be forced to intensively and conservatively censor and filter speech of billions of internet users. Ideas—including ideas with the potential to improve the internet and society—would be abandoned in the face of overwhelming legal risk.**

**We recognize that free speech is not free, and would encourage Congress to focus on legislation that helps us hold bad actors accountable without silencing legitimate users and communities. In addition to our support for the STOP CSAM Act, we support a "bad samaritan" carve-out that would remove 230 protection from legitimately bad actors, such as websites that have the primary purpose of facilitating activities that violate federal criminal law.**

2. Is it your belief that your company should enjoy absolute immunity under Section 230 from suits like *Doe v. Twitter*, No. 21-CV-00485-JCS, 2023 WL 8568911 (N.D. Cal. Dec. 11, 2023), no matter the extent of your company's failure to remove reported child sexual abuse material from the platform or to stop its distribution?

**The company's legal position in the *Doe v. Twitter* matter is set forth in our public filings, and I am not in a position to comment further on the matter, as it is in active litigation. However, I note that the events at issue in that matter all took place years ago under prior management. X also supports the STOP CSAM Act, which would provide for a framework in which to hold companies civilly liable for failure to remove reported child sexual abuse material.**

**Linda Yaccarino – Big Tech and the Online Child Sexual Exploitation Crisis**
**Questions for the Record**
**Submitted February 7, 2024**

## QUESTIONS FROM SENATOR COONS

1. X Corp. ("X") has not published a transparency report with information regarding the United States since April 2023. Why has X not published any transparency report with information regarding the United States in nearly one year?

   **X remains committed to transparency across the company, whether being the first amongst our peers to publish our recommendation algorithm, or making all data related to Community Notes publicly available, as well as open sourcing the code that powers it. Our goal is to return to a regular cadence of publishing global transparency reports in 2024.**

2. Does X measure an estimated total amount of content on the platform that violates its suicide and self-harm policy? If not, why not?

   **Yes. In 2023, more than 900,000 posts and 8,000 accounts were removed for violating our policy around promoting suicide and self harm.**

   a. Does X disclose an estimated total amount of content on its platform that violates its suicide and self-harm policy? If so, please provide a specific citation to where X discloses that information. If not, why not?

      **See answer to Question 1.**

3. X has previously reported how much content it removes under the platform's suicide and self-harm policy.
   a. For content that has been removed, does X measure how many views that content received prior to being removed? If not, why not?

      **X does measure data related to impressions for content posted on the platform.**

   b. For content that has been removed, does X disclose how many views that content received prior to being removed? If so, please provide a specific citation to where X discloses that information. If not, why not?

      **We continue to explore ways to share more context about how we enforce the X Rules. In a previous report, we disclosed data related to impressions of violative content. We will continue to explore the structure of future transparency reports, and we will consider data related to impressions as a reportable metric.**

c. Please provide an estimate of the number of views content that was removed under this policy received in January 2024.

**Data related to January 2024 will be disclosed in a future transparency report.**

d. For content that has been removed, does X measure demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If not, why not?

**X maintains the ability to measure demographic factors about users who view content on the platform.**

e. For content that has been removed, does X disclose demographic factors about users who viewed the violating content, such as how many times the content was viewed by minors? If so, please provide a specific citation to where X discloses that information. If not, why not?

**X does not disclose demographic factors about users who viewed violative content. We will take this under advisement and consideration for future disclosures.**

f. Does X measure the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If not, why not?

**X maintains the ability to measure the number of users that have viewed violative content.**

g. Does X disclose the number of users that have viewed content that was removed under its suicide and self-harm policy multiple times? If so, please provide a specific citation to where X discloses that information. If not, why not?

**X does not disclose data about users who viewed violative content multiple times. We will take this under advisement and consideration for future disclosures.**

4. X utilizes an algorithm to recommend or amplify content to users.
   a. For content that has been removed, does X measure whether and the extent to which the removed content was recommended or amplified by X? If not, why not?

**Restricting the reach of Posts, also known as visibility filtering, is one of our existing enforcement actions that allows us to move beyond the binary "leave up versus take down" approach to content moderation. However, like other social platforms, we have not historically been transparent when we have taken this action. Under certain policies, for example Hateful Conduct, we now add publicly visible labels to Posts identified as potentially violating our policies letting you know we have limited their visibility.**

**These labels bring a new level of transparency to enforcement actions by**

**displaying which policy the Post potentially violates to both the Post author and other users on X. Posts with these labels will be made less discoverable on the platform. Additionally, we will not place ads adjacent to content that we label. You can learn more about the ways we may restrict a Post's reach here. https://help.twitter.com/en/rules-and-policies/enforcement-options**

b. For content that has been removed, does X disclose whether and the extent to which the removed content was recommended or amplified by X? If so, please provide a specific citation to where X discloses that information. If not, why not?

**X does not disclose whether and the extent to which the removed content was recommended. We will take this under advisement and consideration for future disclosures.**

c. For content that has been removed, does X measure how many views the removed content received after having been recommended or amplified? If not, why not?

**X does measure data related to impressions for content posted on the platform.**

d. For content that has been removed, does X disclose the number of views the removed content received after having been amplified or recommended? If so, please provide a specific citation to where X discloses that information. If not, why not?

**We continue to explore ways to share more context about how we enforce the X Rules.**

5. Does X support creating industry-wide transparency requirements to disclose basic safety information, like those included in the *Platform Accountability and Transparency Act*?

**Yes.**

**Linda Yaccarino**
**Chief Executive Officer**
**X Corp.**
**San Francisco, CA**
**Questions for the Record**
**Submitted February 7, 2024**

## QUESTIONS FROM SENATOR BOOKER

1. Trust and safety teams are a vital component in combatting the spread of CSAM, hate speech, violence, and other violative content on tech platforms. Despite this, tech companies have time and time again disinvested from their trust and safety team, especially during changes in leadership.

    a. How has the size of your trust and safety team changed over the past five years? Please provide numbers for each of the past five years.

    **X had 3317 Trust and Safety employees and contractors in May 2022, and 2849 in May 2023. Today, we have approximately 2300 people working on Trust and Safety matters and are building a Trust and Safety Center of Excellence in Austin, Texas, in an effort to bring more agent capacity in-house and rely less on outside contractors. We are currently hiring for full-time agent positions and have a goal of hiring approximately 100 new team members in Austin. A live job posting for this position can be found on our careers page: https://twitter.wd5.myworkdayjobs.com/X/job/Austin-TX/Agent--Trust---Safety--Content-Moderation-_R100044**

    b. Do your trust and safety teams make submissions to the National Center for Missing & Exploited Children's CyberTipline, or is that a separate unit?

    **Yes.**

    c. If it is a separate unit, how many members are on the team and how have those numbers changed over the past five years. Please provide numbers for each of the past five years.

    **N/A**

2. The National Center for Missing & Exploited Children's CyberTipline plays an integral role in combatting child sexual exploitation. The tipline helps law enforcement investigate potential cases and allows prosecutors to bring justice to victims. While federal law requires your company to report to the CyberTipline any apparent violations of federal laws prohibiting child sexual abuse material of which you are aware, there are many gaps.

a. Is there a standard format your reports to the CyberTipline follow? If so, what is that format?

   **Yes, we will follow up with a sample CyberTipline report. Generally, a report contains a complete archive of the account, including content and media, and information such as geolocation, associated emails and/or phone numbers, and IP address(es).**

b. Does your company proactively report planned or imminent offenses?

   **Yes.**

c. Does your company proactively report potential offenses involving coercion or enticement of children?

   **Yes.**

d. Does your company proactively report apparent child sex trafficking?

   **Yes.**

**Questions for the Record from Senator Alex Padilla**
**Senate Judiciary Committee**
**"Big Tech and the Online Child Sexual Exploitation Crisis"**
**Wednesday, January 31, 2024**

## Questions for Linda Yaccarino

1. In your testimony, you shared that less than 1% of X users are below the age of 18.
    a. How many minors are registered X account holders?

    **As of January 2024, in the United States, there are approximately 650,000 daily active users between the ages of 13-17.**

    b. Does X plan to provide guidance to them and their caregivers about digital online health and safety?

    **X provides all users and the public information on safety and security on X via our Help Center. https://help.twitter.com/en/safety-and-security**

2. In your testimony, you shared that users between the ages of 13 and 17 are automatically assigned to a private default setting and they cannot accept a message from anyone they do not approve.
    a. How are you ensuring that the burden is not on young people to make adult-level decisions about safety on the services that you operate?

    **By implementing privacy and safety by default, X makes it simple for minors to use X safely.**

    **Accounts belonging to known minors will be defaulted to a "protected" setting. This means that known minors will receive a request when new people want to follow them (which they can approve or deny), that their posts will only be visible to their followers, and that their posts will only be searchable by them and their followers (i.e. they will not appear in public searches). Under this setting, accounts belonging to known minors will be restricted to receiving DMs from accounts they follow by default. We also utilize an age lock. Once a new user enters a date of birth that makes them under the age of 18, they will be stopped from re-entering a new date of birth for that account.**

    **We also take steps to limit exposure to sensitive content. Known minors or viewers who do not include a birth date on their profile are restricted from viewing specific forms of sensitive media such as adult content. X obscures sensitive media behind notices and interstitials. This includes our product age restrictions that restricts known minors from viewing adult content.**

**In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.**

**More information on our protected account settings can be found in our Help Center.**
**https://help.twitter.com/en/safety-and-security/public-and-protected-posts**

b. In the last 4 years, how often has the company blocked products from launching because they were not safe enough for minors, or withdrawn products from the market after receiving feedback on the harms they were causing?

**To the best of my knowledge, since I joined the company X has not launched products targeted at minors, nor has X withdrawn products from the market due to harms to minors.**

c. Since Mr. Musk took ownership of the company, how often has the company blocked products from launching because they were not safe enough for minors, or withdrawn products from the market after receiving feedback on the harms they were causing?

**To the best of my knowledge, since I joined the company X has not launched products targeted at minors, nor has X withdrawn products from the market due to harms to minors.**

3. Existing detection tools for keeping child sexual abuse material from spreading online rely on hashed images of already identified CSAM imagery. There are tools like PhotoDNA and Google's CSAI match tool available for identifying this content. A challenge I hear raised frequently is identifying and removing novel images that have not already been hashed.
   a. What would it take to develop better technology to accurately identify and limit the spread of novel CSAM images?

**At X we use a combination of commercial, proprietary, and open-source technology to identify, hash, report, and remove novel CSAM images, and we will continue to enhance this technology while partnering with advanced technology providers like Thorn.**

**X is defending itself in a class action lawsuit related to its use of one of these tools. In *Martell v. X Corp.*, Civil No. 05449 (N.D. Ill. 2023), a private plaintiff has alleged that X's use of PhotoDNA to combat CSAM violates Illinois' Biometric Information Privacy Act. Meritless lawsuits such as *Martell* discourage critical innovation in this space. We encourage Congress to explore federal protections for the development and use of anti-CSAM technologies.**

b. Are there interventions from Congress that would facilitate identification of CSAM?

**Investment in research and development of advanced technologies to detect CSAM would benefit the internet ecosystem.**

c. Based on your company's experience trying to address online sexual exploitation and abuse of minors, are there areas where Congress could be helpful in tackling this problem?

**Increased investment in law enforcement capabilities to investigate and prosecute criminals that traffic CSAM would strengthen the ecosystem as a whole. Increased investment and resources for the National Center for Missing and Exploited Children to develop their technical capabilities would accelerate the investigation and prosecution of criminals around the world.**

4. AI models are making it easier to develop synthetic CSAM. These are either altered images of real people, or wholly synthetic individuals. Policymakers are grappling with what this will mean for law enforcement efforts to hold perpetrators accountable and identify children who are being harmed. In addition to processing a higher volume of Cybertips, investigators will have the added challenge of determining whether the victim in the scenario is in fact a real person. And cases are already being reported where AI generative technologies are being employed to facilitate the grooming and sextortion of minor victims.
    a. What are you doing to identify and remove AI-generated CSAM on your services?

    **X utilizes a combination of technology and human capacity to enforce our Child Sexual Exploitation (CSE) policy. This policy covers media, text, illustrated, or computer-generated images.**

    b. Do you flag for NCMEC if you perceive the CSAM to be AI-generated?

    **Currently, we do not include this in our NCMEC reports. However, we are evaluating the requirements to include this new field in our reports.**

    c. How prevalent is this kind of content?

    **We do not currently have prevalence data on this type of content.**

    d. How do you anticipate the rise of AI-generated CSAM will impact NCMEC's ability to process and refer Cybertips to law enforcement?

    **We defer to NCMEC's expertise and experience on the impact on their ability to process and refer CyberTips to law enforcement, however, we believe that individuals who traffic computer-generated CSAM should be**

**investigated and prosecuted.**

e. Recently, A.I.–generated explicit images of a major pop superstar were distributed widely online without her consent. That story drew attention to a growing problem over the last year facilitated by AI tools: the generation of deepfake, nonconsensual, sexually explicit imagery of everyday people, including our young people. Will you commit to reporting on the prevalence of this new problem and the steps your company is taking to address this horrendous abuse?

**Non-consensual nudity (NCN) has no place on X and we will continue to rigorously enforce this policy and improve our approach. As the challenge of AI-generated NCN evolves, we will commit to working with all stakeholders to meet the challenge. We support Congressional efforts to criminalize the distribution of "deepfake" non-consensual intimate imagery.**

f. Are there technical or legal barriers that your company has identified preventing thorough redteaming of AI models to ensure they do not generate CSAM?

**X does not develop AI models that generate images.**

5. How companies choose to allocate their resources illustrates their true priorities.
   a. What percentage of your company's budget is dedicated to addressing child safety on your platform?

   **Child safety is embedded into our company values and operations; there is not a specific budget allocation.**

   b. What process or assessment of risk on the platform informed that figure?

   **See answer above.**

   c. How many layers of leadership separates your trust and safety leaders from you and Mr. Musk?

   **Our Trust & Safety leadership report directly to the CEO.**

   d. At X, which member of the leadership team ultimately approves product decisions that impact user safety?

   **X's leadership across the company considers user safety in their decision-making, including product decisions.**

6. The companies represented at the hearing have the money and resources to hire teams of Trust & Safety professionals and build bespoke tools to aid with content moderation and integrity work as well as the detection of content like CSAM on their services. This is not

necessarily the case for the rest of the tech sector. These are industry-wide problems and will demand industry-wide professionalization and work.

a. What is X currently doing to support access to open-source trust & safety tools for the broader tech ecosystem?

**Currently, we utilize a mix of proprietary tools and licensed technologies, for example Thorn's Safer tool. (https://safer.io/) We recognize the power of open source technology and if there are technologies that can be open sourced without compromising operational integrity and preventing bad actors from exploiting the technology, we will support those efforts.**

b. And if X is not doing anything now, will you commit to supporting the development of these kinds of resources?

**X is committed to technology transparency, as evidenced by the publishing of our recommendation algorithm, currently available on GitHub. https://github.com/twitter/the-algorithm. The technology powering our Community Notes product is also open source, and also available on GitHub. https://github.com/twitter/communitynotes**

7. One necessary element of keeping our kids safe is preventing harms in the first place. The National Center for Missing and Exploited Children partnered with the White House, the Department of Justice, and the Department of Homeland Security to create "The Safety Pledge" initiative to combat online child exploitation in September 2020. I understand more government backed public awareness campaigns are being developed.

a. Are you partnering with the federal government to distribute health and safety resources to young people?

**We support the NCMEC with advertising credits to promote their campaigns on X (via their accounts @MissingKids and @NetSmartz).**

b. What are you proactively doing to educate the minors that use your services about online health and safety?

**We provide advertising credits to organizations around the world working on digital safety.**

8. Sextortion has become increasingly prevalent. Offenders may use grooming techniques or basic trickery to manipulate victims into providing nude or partially nude images of themselves, which are then used to coerce victims into sending more graphic images and videos or pay a ransom. These criminals often threaten to post the images or sensitive images publicly or send them to the victim's friends and family if the child does not comply. From May 2022 to October 2022, U.S. law enforcement and NCMEC witnessed an alarming increase in CyberTips and reports where minors have been sextorted for money. Many young boys, including in California, have committed suicide out of desperation, leaving their loved ones devastated.

a. How is your company responding to the growing threat of financial sextortion?

**We use a mixture of proactive detection, user reporting, and human capacity to enforce our private information and media policy. We are currently beta testing a text-based machine learning classifier developed by Thorn that detects sextortion, CSAM, child-access, self-generated CSAM, and child abuse discussion.**

**An important component in combating extortion is our private information policy. Under this policy, you may not publish or post other people's private information without their express authorization and permission. We also prohibit threatening to expose private information or incentivizing others to do so.**

**In addition, you may not share private media, such as images or videos of private individuals, without their consent. However, we recognise that there are instances where users may share images or videos of private individuals, who are not public figures, as part of a newsworthy event or to further public discourse on issues or events of public interest. In such cases, we may allow the media to remain on the platform.**

**Sharing someone's private information online without their permission, sometimes called doxxing, is a breach of their privacy and of the X Rules. Sharing private information can pose serious safety and security risks for those affected and can lead to physical, emotional, and financial hardship.**

**When reviewing reports under this policy, we consider a number of things, including:**

- ***What type of information is being shared***
    - **We take this into consideration because certain types of private or live information carry higher risks than others, if they're shared without permission. Our primary aim is to protect individuals from potential physical harm as a result of their information being shared, so we consider information such as physical location and phone numbers to be a higher risk than other types of information. We define "live" as real-time and/or same-day information where there is potential that the individual could still be at the named location.**
- ***Who is sharing the information***
    - **We also consider who is sharing the reported information and whether or not they have the consent of the person it belongs to. We do this because we know that there are times when people may want some forms of their personal information to be shared publicly. For example, sharing a personal phone number or email for professional networking or to coordinate**

**social events or publicly sharing someone's home addresses or live locations to seek help after a natural disaster.**

- ***Whether the information available elsewhere online***
  - **If the reported information was shared somewhere else before it was shared on X, e.g., someone sharing their personal phone number on their own publicly accessible website, we may not treat this information as private, as the owner has made it publicly available. Note: we may take action against home addresses being shared, even if they are publicly available, due to the potential for physical harm.**
- ***Why the information is being shared***
  - **We also factor in the intent of the person sharing the information. For example, if we believe that someone is sharing information with an abusive intent, or to harass or encourage others to harass another person, we will take action. On the other hand, if someone is sharing information in an effort to help someone involved in a crisis situation like in the aftermath of a violent event, we may not take action. Note: regardless of intent, if the information is not shared during a crisis situation to assist with humanitarian efforts or in relation to public engagement events, we will remove any posts or accounts that share someone's live location.**
- ***Sharing private media***
  - **Posting images is an important part of our users' experience on X. Where individuals have a reasonable expectation of privacy in an individual piece of media, we believe they should be able to determine whether or not it is shared. Sharing such media could potentially violate users' privacy and may lead to emotional or physical harm. When we are notified by individuals depicted, or their authorized representative, that they did not consent to having media shared, we will remove the media. This policy is not applicable to public figures.**

**Under this policy, you can't share the following types of private information, without the permission of the person who it belongs to:**

- **home address or physical location information, including street addresses, GPS coordinates or other identifying information related to locations that are considered private;**
- **live location information, including information shared on X directly or links to 3rd-party URL(s) of travel routes, actual physical location, or other identifying information that would reveal a person's location, regardless if this information is publicly available;**
- **identity documents, including government-issued IDs and social security or other national identity numbers – note: we may make**

**limited exceptions in regions where this information is not considered to be private;**
- **contact information, including non-public personal phone numbers or email addresses;**
- **financial account information, including bank account and credit card details;**
- **other private information, including biometric data or medical records;**
- **media of private individuals without the permission of the person(s) depicted; and**
- **media depicting prisoners of war posted by government or state-affiliated media accounts on or after April 5, 2022.**

**The following behaviors are also not permitted:**
- **threatening to publicly expose someone's private information;**
- **sharing information that would enable individuals to hack or gain access to someone's private information without their consent, e.g., sharing sign-in credentials for online banking services;**
- **asking for or offering a bounty or financial reward in exchange for posting someone's private information;**
- **asking for a bounty or financial reward in exchange for not posting someone's private information, sometimes referred to as blackmail.**

b. What methods are in place to detect and disrupt this type of abuse in real time?

**See above answer.**

c. What kind of user education and awareness are you engaged in?

**We maintain a comprehensive Help Center and regularly post on our @Safety account on X.**

d. Are you aware of a higher prevalence of sexual extortion or abuse against certain demographics among young users? If not, will you commit to studying this issue and making that kind of information available to improve public education and protection measures?

**We will continue to stay educated on the latest trends in exploitative behavior on our platform and across industry. Many cross-sector groups and convenings, whether NCMEC's CyberTipline Roundtable or Virtual Global Taskforce meeting, the INHOPE conference, or Tech Coalition member events, provide a great opportunity for knowledge sharing on the latest threats.**

9. Young people need to be at the center of regulatory discussions, and they need to be at the table as products and services they use are designed.

a. Are you engaging young adults and youth in your conversations and policies around Trust and Safety on the platform?

**We are constantly gathering input and feedback from our users around the world on issues of trust and safety.**

b. How do you proactively keep up to speed with the most pressing issues facing young people online?

**As the global town square of the internet, every day there is robust conversation about online safety on X, from discussion of research, to announcements about technologies or products from a range of companies, to in-depth reporting about the challenges that the internet ecosystem faces. X is the place where experts, policymakers, and parents come to discuss the most pressing issues facing young people online, and where youth activists advocate for change.**

10. While employment figures alone do not reflect a company's ability to address trust & safety on their services, the precipitous decline in the trust and safety expertise and personnel at X has been alarming. According to disclosures X made to Australian regulators, between October 28, 2022, and May 31, 2023, your trust and safety staff globally had been reduced from 4,062 to 2,849 employees and contractors. Engineers focused on trust and safety issues at X had been reduced from 279 globally to 55. Full-time employee content moderators had been reduced from 107 to 51. Content moderators employed on contract fell from 2,613 to 2,305.
    a. Does X currently have a specific person in charge of Trust & Safety strategy and policy decisions?

    **Yes.**

    b. Who does that person directly report to?

    **The CEO.**

    c. How many employees at X are focused on strategic Trust & Safety matters?

    **We have approximately 2300 people dedicated to Trust & Safety matters.**

    d. What resources does X dedicate to child safety?

    **Child safety is embedded into our company values and across all of our operations; there is not a specific budget or resource allocation.**

    e. How does X's current child safety work compare financially, technically, and personnel-wise to two years ago?

**An important shift and enhancement in our ability to send reports to the CyberTipline was the ten-fold increase and investment we made in training our agents in CyberTipline reporting. Prior to acquisition, we had approximately 20 agents that were trained and authorized to report to NCMEC. Now, we have approximately 200 agents that are authorized and trained and sending reports to the CyberTipline. This increase has led to an average of 64,000 manual reports per month sent to the CyberTipline, up from an average of 6,300 manual reports per month.**

11. A Wall Street Journal investigation last year with the Stanford Internet Observatory and UMass Rescue Lab identified a network of more than 500 accounts of young users advertising their self-generated illicit sexual media on social media, especially Instagram and X, with tens of thousands of likely buyers. X insisted that they took action to address failures. However, months later, Stanford researchers found the problem persisted.
    a. The Stanford researchers were not able to do a complete reassessment of X because the company removed access to its Academic API offerings. Why did X retire these offerings and will you reevaluate the posture of the company with respect to academic researchers?

    **We are re-evaluating our academic research program and look forward to sharing more soon.**

**Senator Peter Welch**
**Senate Judiciary Committee**
**Written Questions for Linda Yaccarino**
**Hearing on "Big Tech and the Online Child Sexual Exploitation Crisis"**
**January 31, 2024**

X recently announced that the company plans to build a Trust and Safety Center for Excellence in Austin, Texas and will bring on 100 new employees. This announcement comes after X carried out a series of layoffs in 2023, which included cuts to X's Trust & Safety team.

1. How many people were laid off from X's Trust & Safety team in 2023?

   **The Trust & Safety organization has been the least impacted by reductions in force. Since acquisition, we have right-sized the company and are hiring across departments, including Trust & Safety.**

2. Why did you decide to add their jobs back? If the 100 new employees will have substantially different duties than the previously laid off employees, please describe the differences.

   **Our Trust and Safety Center of Excellence will consist of a mix of agents and policy specialists, who work on a range of content moderation issues and enforcement.**

3. Even before the layoffs in 2023, tech companies like X lacked appropriate policies to counter dangerous content on their platforms. How do you plan to address these issues moving forward?

   **We will continue to improve the safety of the platform and our users, and we continue to conduct regular reviews of our policies, enforcement guidance, and training materials as we learn more about new content and behavioral challenges.**

4. How much of your content moderation is managed by artificial intelligence?

   **We use a mix of technology and human capacity to enforce our policies. Content moderation is not managed by artificial intelligence, rather by members of our Trust & Safety team.**

5. Is it your view that artificial intelligence can replace human judgment in identifying and removing false or harmful content? If not, when is human judgment necessary?

   **We should continue to utilize a mix of technology and human capacity to enforce our policies and maintain the integrity of our platform.**

6. How have your Trust & Safety teams been trained on how to handle false or illegal AI-generated content?

**X has a clear policy on the use of synthetic and manipulated media. This policy expressly provides that users "may not share synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm ('misleading media')." As part of this policy, which extends to all advertisements, X may label posts containing misleading media to help people understand their authenticity and to provide additional context.**

**In order for content with misleading media (including images, videos, audios, gifs, and URLs hosting relevant content) to be labeled or removed under this policy, it must:**

- **Include media that is significantly and deceptively altered, manipulated, or fabricated, or**
- **Include media that is shared in a deceptive manner or with false context, and**
- **Include media likely to result in widespread confusion on public issues, impact public safety, or cause serious harm**

**We use the following criteria as we consider posts and media for labeling or removal under this policy as part of our ongoing work to enforce our rules and ensure healthy and safe conversations on X:**

> ***1. Is the content significantly and deceptively altered, manipulated, or fabricated?***
>
> **In order for content to be labeled or removed under this policy, we must have reason to believe that media are significantly and deceptively altered, manipulated, or fabricated. Synthetic and manipulated media take many different forms and people can employ a wide range of technologies to produce these media. Some of the factors we consider include:**
>
> - **whether media have been substantially edited or post-processed in a manner that fundamentally alters their composition, sequence, timing, or framing and distorts their meaning;**
> - **whether there are any visual or auditory information (such as new video frames, overdubbed audio, or modified subtitles) that has been added, edited, or removed that fundamentally changes the understanding, meaning, or context of the media;**
> - **whether media have been created, edited, or post-processed with enhancements or use of filters that fundamentally changes the understanding, meaning, or context of the content; and**
> - **whether media depicting a real person have been fabricated or simulated, especially through use of artificial intelligence algorithms**
>
> **We will not take action to label or remove media that have been edited in ways that do not fundamentally alter their meaning, such as retouched photos or color-corrected videos.**
>
> **In order to determine if media have been significantly and deceptively**

**altered or fabricated, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been altered or fabricated, we may not take action to label or remove them.**

***2. Is the content shared in a deceptive manner or with false context?***

**We also consider whether the context in which media are shared could result in confusion or suggests a deliberate intent to deceive people about the nature or origin of the content, for example, by falsely claiming that it depicts reality. We assess the context provided alongside media to see whether it provides true and factual information. Some of the types of context we assess in order to make this determination include:**

- **whether inauthentic, fictional, or produced media are presented or being endorsed as fact or reality, including produced or staged works, reenactments, or exhibitions portrayed as actual events;**
- **whether media are presented with false or misleading context surrounding the source, location, time, or authenticity of the media;**
- **whether media are presented with false or misleading context surrounding the identity of the individuals or entities visually depicted in the media;**
- **whether media are presented with misstatements or misquotations of what is being said or presented with fabricated claims of fact of what is being depicted**

**We will not take action to label or remove media that have been shared with commentary or opinions that do not advance or present a misleading claim on the context of the media such as those listed above.**

**In order to determine if media have been shared in a deceptive manner or with false context, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been shared with false context, we will not label or remove the content.**

***3. Is the content likely to result in widespread confusion on public issues, impact public safety, or cause serious harm?***

**Posts that share misleading media are subject to removal under this policy if they are likely to cause serious harm. Some specific harms we consider include: Threats to physical safety of a person or group; incitement of abusive behavior to a person or group; risk of mass violence or widespread civil unrest; risk of impeding or complicating provision of public services, protection efforts, or emergency response. We also consider threats to the privacy or to the ability of a person or group to freely express themselves or**

**participate in civic events, such as: stalking or unwanted and obsessive attention; targeted content that aims to harass, intimidate, or silence someone else's voice; or voter suppression or intimidation. We also consider the time frame within which the content may be likely to impact public safety or cause serious harm, and are more likely to remove content under this policy if immediate harm is likely to result.**

**Posts with misleading media that are not likely to result in immediate harm but still have a potential to impact public safety, result in harm, or cause widespread confusion towards a public issue (health, environment, safety, human rights and equality, immigration, and social and political stability) may be labeled to reduce their spread and to provide additional context.**

**While we have other rules also intended to address these forms of harm, including our policies on violent threats, civic integrity, and hateful conduct, we will err toward removal in borderline cases that might otherwise not violate existing rules for Posts that include misleading media.**

**The consequences for violating our synthetic and manipulated media policy depends on the severity of the violation. For high-severity violations of the policy, including misleading media that have a serious risk of harm to individuals or communities, we will require you to remove this content. In circumstances where we do not remove content which violates this policy, we may provide additional context on posts sharing the misleading media where they appear on X. This means we may: Apply a label and/or warning message to the post; show a warning to people before they share or like the post; reduce the visibility of the post on the platform and/or prevent it from being recommended; turn off likes, replies, and Reposts; and/or provide a link to additional explanations or clarifications, such as relevant X policies. In most cases, we will take a combination of the above actions on posts we label.**

**If we determine that an account has advanced or continuously shares harmful misleading narratives that violate the synthetic and manipulated media policy, we may temporarily reduce the visibility of the account or lock or suspend the account. If users believe that their account was locked or suspended in error, they can submit an appeal.**

7. How does X plan on addressing the large amount of disinformation that could be spread on its platform during the 2024 election?

**X's purpose is to serve the public conversation, as people from around the world come together in an open and free exchange of ideas. Our team has and will continue to actively work to protect the integrity of the public conversation, by ensuring that users have access to real-time information and safeguarding the platform for everyone.**

**The public conversation occurring on X is never more important than during elections and other civic events. Any attempt to undermine the integrity of our service is antithetical to our fundamental rights and undermines the core tenets of freedom of expression, the value upon which our company is based.**

**We believe we have a responsibility to protect the integrity of all conversations from interference and manipulation. We prohibit attempts to use our services to manipulate or disrupt civic processes, including through the distribution of false or misleading information about the procedures or circumstances around participation in a civic process. This includes posting or sharing content that may suppress participation or mislead people about when, where, or how to participate in a civic process.**

**Beyond our policies and enforcement, we believe that X users play a pivotal role in helping provide needed context and accurate information about content that may be misleading or false.**

**Product Innovation: Community Notes**

**Community Notes, which focuses on offering context and surfacing credible information, represents a fundamental shift in how X mitigates mis- and disinformation. Community Notes aims to create a better-informed world by empowering X users to collaboratively add helpful notes to posts that might be misleading. Contributors can leave notes on any post and if enough contributors from different points of view rate that note as helpful, the note will be publicly shown the post. We believe that Community Notes is an inherently scalable and localized response to the challenge of disinformation. By making this feature an integral and highly visible part of X, and by ensuring that the user interface is simple and intuitive, we are investing in a tool that can be truly global in its application. It also reduces our reliance on forms of content moderation that are more centralized, manual and bespoke; or which require intensive and time-consuming interactions with third parties.**

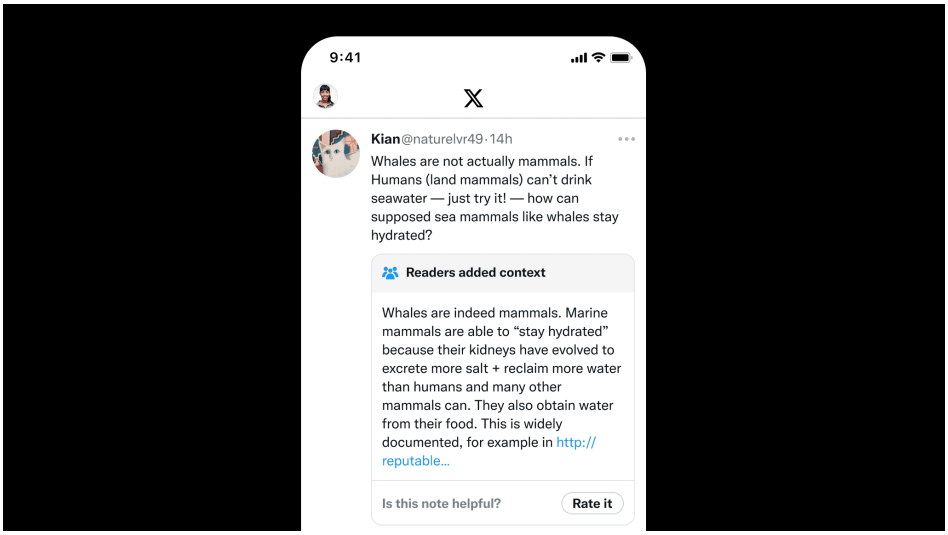
**Here is how it works:**

- **Contributors write and rate notes: Contributors are people on X who sign up to write and rate notes. The more people that participate, the better the program becomes.**
- **Only notes rated helpful by people from diverse perspectives appear on posts: Community Notes do not work by majority rules. To identify notes that are helpful to a wide range of people, notes require agreement between contributors who have sometimes disagreed in their past ratings. This helps prevent one-sided ratings. We have published and will continue to learn more about how Community Notes handles diverse perspectives.**
- **X does not choose what shows up, the people do: X does not write, rate, or moderate notes (unless they break the X Rules.) We believe giving people a voice to make these choices together is a fair and effective way to add**

**information that helps people stay better informed.**

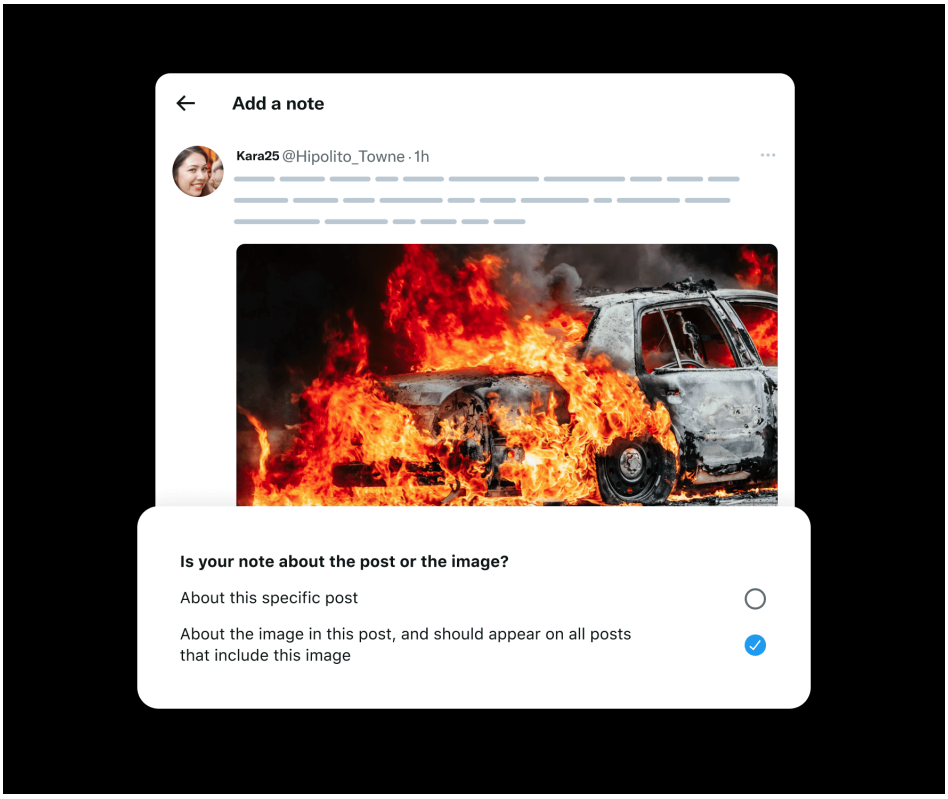
- **Open-source and transparent: It is important for people to understand how Community Notes work to be able to help shape it.**
- **The program is built on transparency: all contributions are published daily, and our ranking algorithm can be inspected by anyone. Learn more about how it works through our dedicated Community Notes Guide.**

**We acknowledge and are keenly aware that a product like this can be subject to attempts of abuse and manipulation, which we proactively assess. You can read more here on how we are thinking about quality control, guardrails, circuit breakers, and the various remediations we have in place to challenge bad actors.**




*Notes on Media*

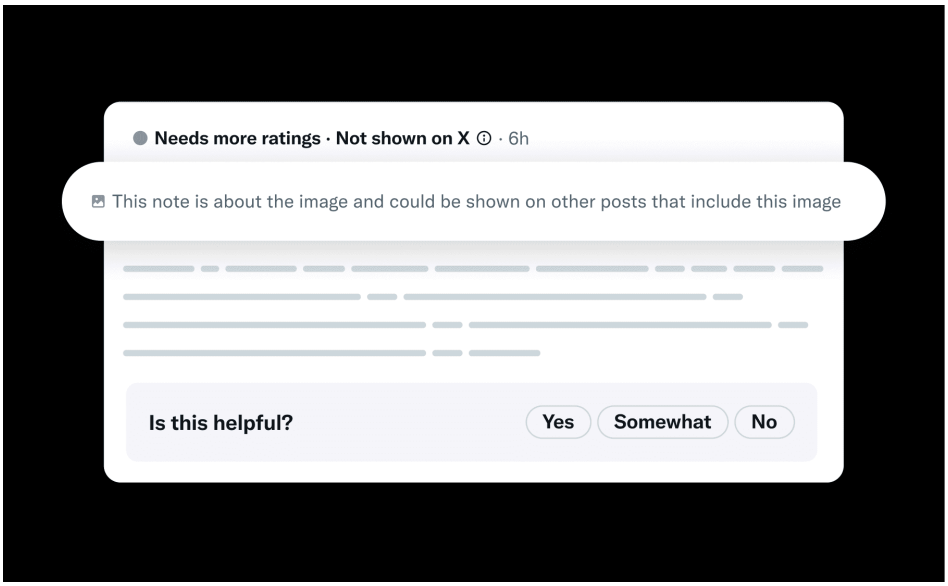
**Community Notes are frequently added to posts that feature images or videos. In many cases, these notes can provide valuable context, not just for a single post, but for any post containing the same media. This feature is especially important for addressing the challenges of media produced by generative artificial intelligence tools. Contributors with a Writing Impact score of 10 or above have the option to write notes about the media found within posts, as opposed to focusing on the specific post. Contributors should select this option when they believe the context added would be helpful independently of the post the note is attached to.**




**Tagging notes as “about the image” makes them visible on all posts that our system identifies as containing the same image. These notes, when deemed Helpful, accumulate view counts from all the posts they appear in, but only count as one Writing and Rating Impact for the author and raters.**

**When someone rates a media note, the rating is associated with the post on which the note appeared. This allows Community Notes to identify cases where a note may not apply to a specific post.**

**For example, because of this new feature, notes related to the Israel-Hamas conflict have been displayed on 10,000+ posts. This number grows automatically if the relevant images and video are re-used in new posts.**




**Currently, this feature is experimental and only supports posts with a single image. We are actively working on expanding it to support posts with multiple images, GIFs, and videos. Stay tuned for updates.**

***Effectiveness & Research***

**We already know that the Community Notes feature is effective. According to the results of four surveys run at different times between August, 2021 and August, 2022, a person who sees a Community Note is, on average, 20-40% less likely to agree with the substance of a potentially misleading post than someone who sees the post alone. Survey participation ranged from 3,000 to more than 19,000 participants, and the results were consistent throughout the course of the year, even as news and post topics changed. We also see that Community Notes informs sharing behavior. Analyzing our internal data, we have found that a person on X who sees a note is, on average, 15-35% less likely to like or repost a post than someone who sees the post alone. In our most recent survey, Community Notes were found to be informative regardless of a person's self-identified political party affiliation — there was no statistically significant difference in average informativeness across party identification.**

**We published a research paper on Community Notes that you can read here. It goes into more detail on how we have been measuring efficacy. In addition, all Community Notes contributions are publicly available on the Download Data page of the Community Notes site so that anyone has free access to analyze the data, identify problems, and spot opportunities to make the product better.**

**We know there are many challenges involved in building an open, participatory system like Community Notes — from making it resistant to manipulation attempts, to ensuring it is not dominated by a simple majority or biased because of the distribution of contributors.**

**We have been building Community Notes (formerly called Birdwatch) in public since January 2021, and have intentionally designed it to mitigate potential risks. We have seen encouraging results, but we are constantly designing for challenges that could arise. Here are a handful of particular challenges we are aware of as well as steps we are taking to address them:**

***Preventing Coordinated Manipulation Attempts***

**Attempts at coordinated manipulation represent a crucial risk for open rating systems. We expect such attempts to occur, and for Community Notes to be effective, it needs to be resistant to them. The program currently takes multiple steps to reduce the potential for this type of manipulation:**

- **First, all X accounts must meet the eligibility criteria to become a Community Notes contributor. For example, having a unique, verified phone number. These criteria are designed to help prevent the creation of large numbers of fake or sock puppet contributor accounts that could be used for inauthentic rating.**
- **Second, Community Notes does not work like many engagement-based ranking systems, where popular content gains the most visibility and people can coordinate to mass upvote or downvote content they do not like or agree with. Instead, Community Notes uses a bridging algorithm — for a note to be shown on a post, it needs to be found helpful by people who have tended to disagree in their past ratings. Academic research indicates that bridging-based ranking can help to identify content that is healthier and higher quality, and reduce the risk of elevating polarizing content.**
- **In addition to requiring ratings from a diversity of contributors, Community Notes has a reputation system in which contributors earn helpfulness scores for contributions that people from a wide range of perspectives find helpful. Helpfulness scores give more influence to people with a track record of making high-quality contributions to Community Notes, and lower influence to new accounts that have yet to demonstrate a track record of helpful ratings and contributions.**
- **Lastly, Community Notes tracks metrics that alert the team if suspicious activity is detected, and has a set of guardrails and procedures to identify if contribution quality falls below set thresholds. This helps Community Notes to proactively detect potential coordination attempts and impacts to note quality.**

***Reflecting Diverse Perspectives, Avoiding Biased Outcomes***

**Community Notes will be most effective if the context it produces can be found to be helpful by people of multiple points of view and not just people from one group or another. To work towards this goal, Community Notes currently takes the following steps:**

- **First, as described above, Community Notes uses a bridging based algorithm**

**to identify notes that are likely to be helpful to people from many points of view. This helps to prevent one-sided ratings and to prevent a single group from being able to engage in mass voting to determine what notes are shown.**

- **Second, Community Notes can proactively seek ratings from contributors who are likely to provide a different perspective based on their rating history. This is currently done in the Needs Your Help tab, and we are exploring new ways to quickly collect ratings on notes from a wide range of contributors.**
- **Third, to help ensure that people of diverse backgrounds and viewpoints feel safe and empowered to participate, Community Notes has implemented program aliases that are not publicly associated with contributors' X accounts. This can help prevent one-sidedness by providing more diverse contributors with a voice in the system.**
- **Finally, we regularly survey representative samples of X customers who are not Community Notes contributors to assess whether a broad range of people on X are likely to find the context in Community Notes to be helpful, and whether the notes can be informative to people of different points of view. This is one indicator of Community Notes' ability to be of value to people from a wide range of perspectives vs. to be biased towards one group or viewpoint. X customers who are not enrolled Community Notes contributors can also provide rating feedback on notes they see on X. This provides an additional indicator of note helpfulness observed over time.**

**Expansion & Localization**

**Community Notes are now publicly visible to everyone on X. Users in 65 countries, including the US, the UK, Japan, Brazil, Mexico, Ireland, Canada, Spain, Portugal, Italy, Germany, Austria, Belgium, France, Switzerland, Luxemburg, Netherlands, Australia, New Zealand, Slovakia, Algeria, Bahrain, Egypt, Israel, Jordan, Kuwait, Lebanon, Morocco, Oman, Palestinian Territories, Qatar, Tunisia, United Arab Emirates, and just recently, Hong Kong, South Korea, and Taiwan, can now contribute to the program. Over the coming months, users in more markets will be able to contribute notes and the product will be localized further. We currently have over 400,000 users enrolled in Community Notes.**

**Over time, users writing in any language should be able to contribute to Community Notes and the most helpful contributions will be surfaced to inform readers. Eventually, we can see a future where attempts to spread disinformation are consistently flagged by conscientious users seeking to share important context and facts with citations.**

**The technology-first strategy evidenced by Community Notes is reflective of how we intend to approach content moderation going forward. We believe that this approach has obvious advantages over more centralized methods of content moderation, which have always faced the same two challenges: speed and scale.**

**This is an open and transparent process. That is why we have made the Community**

**Notes algorithm open source and publicly available on GitHub, along with the data that powers it so anyone can audit, analyze or suggest improvements.**

X recently instituted full end-to-end encryption for some direct messages on its platform. All questions below pertain to X's use of end-to-end encryption.

8. Please describe why you chose to implement end-to-end encryption for direct messages and the benefits you see from it.

   **At this time, X does not offer full end-to-end encryption. Encrypted Direct Messages are limited to only subscribers of X Premium, and only text is encrypted.**

   **Users need to satisfy the following conditions in order to send and receive encrypted messages:**
   a. **both sender and recipient are on the latest X apps (iOS, Android, Web);**
   b. **both sender and recipient are verified users or affiliates to a verified organization; and**
   c. **the recipient follows the sender, or has sent a message to sender previously, or has accepted a Direct Message request from the sender before.**

9. How do you balance the benefits of encryption with the need for law enforcement to be able to track down wrongdoers on your platform?

   **We limit the availability of encryption to X Premium subscribers, which reduces the number of accounts eligible to use encryption, while also providing more personal and financial information that law enforcement could utilize.**

10. Given X's recent implementation of end-to-end encryption, please explain the steps and processes you use to identify child sexual abuse material, how you remove it, and how you report it to law enforcement.

    **By encrypting only text, we are able to scan Direct Messages for known or potential CSAM.**

    **Anyone can report potential CSAM, whether they have an X account or not. In the majority of cases, the consequence for violating our CSE and CSAM policy is immediate and permanent suspension from the platform. In addition, violators will be prohibited from creating any new accounts in the future. When we are made aware of content depicting or promoting child sexual exploitation, including links to third party sites where this content can be accessed, we immediately remove it without further notice and report to the National Center for Missing & Exploited Children (NCMEC).**

    **Our proactive detection efforts are primarily driven by the following tools:**
    - **Hash-sharing: Our current methods of surfacing potentially violating content for human review include leveraging the hashes provided by NCMEC and**

**industry partners - which makes this the most widely used form of CSAM detection. We scan all media uploaded to X for matches to hashes of known CSAM sourced from NGOs, law enforcement, and other platforms. Users posting known content are immediately permanently suspended and reported to NCMEC. For videos we use a proprietary hashing algorithm produced by Thorn.**

- **Automatic text detection: We have a variety of tools to access the likelihood that a post is advertising or promoting the sharing of child sexual abuse material. Some of these defenses lead to automatic suspensions while other users are flagged for human review.**
- **For videos we use a proprietary hashing algorithm produced by Thorn.**
- **PhotoDNA and internal proprietary tools: a combination of technology solutions are used to surface accounts violating our rules on Child Sexual Exploitation.**
- **Media Risk Scanning: We receive a media classifier score through Safer and it is used to filter false positive hash matches at the moment. We use a novel classifier model to rate media shared through posts' likelihood of being CSAM. Media that receives a high score is then sent to human review.**

**Additionally, any attempts to circumvent an enforcement action (such as a permanent suspension) by creating additional accounts or repurposing existing accounts to replace or mimic a suspended account are considered a violation of our ban evasion policy and it will result in permanent suspension at first detection.**

**Post-Hearing Questions for the Record**
**Submitted to X CEO Linda Yaccarino**
**From Senator Laphonza Butler**

**"Protecting Our Children Online: Big Tech and the Crisis of Online Child Sexual Exploitation"**

**January 31, 2024**

1. **Family and parental control tools:** I was glad to hear that you have spent time talking with parents and what their families need from your products. I was also glad to hear your companies have a Family Center, or other similar tools, to give parents more insight and control over how their children are using your platforms and apps.

   a. How do you advertise this feature to parents?

      **We do not currently have parental tools or a Family Center.**

   b. Can you share data on how many Family Center/parental tools users there are in proportion to total minors on your platforms and products?

      **N/A**

**a.**

**Questions for Linda Yaccarino (X) - Grassley**

*Please answer each question to the fullest possible extent. If your platform is unable to answer a particular question or does not have requested data, explain why. Each question refers to your company in addition to any corporate affiliates, including parent and subsidiary companies.*

1. Current law requires that a provider of a report of suspected CSAM to the National Center for Missing and Exploited Children's (NCMEC) CyberTipline preserve "any visual depictions, data, or other digital files that are reasonably accessible and may provide context or additional information about the reported material or person" for a minimum of 90 days. 18 U.S.C. 2258A(h)(1-2). The recent explosion of suspected abuse has presented unprecedented challenges for law enforcement to follow up on leads before companies discard or delete essential data and information. There is nothing preventing tech companies from preserving relevant material beyond the statutorily-mandated 90-day period.
    a. How long does X voluntarily preserve and retain data contained in and related to its reports to the CyberTipline?

       **X retains the associated account data for 90 days, unless it is subject to a legal obligation to preserve the data for a longer period. However, this data is provided to NCMEC who may then retain the data for investigative purposes in line with their own retention policies.**

    b. The massive influx of reports to the CyberTipline naturally results in law enforcement entities having to conduct and finish investigations beyond 90 days of an initial report to the CyberTipline. Retaining relevant information for longer periods could significantly advance law enforcement's ability to thoroughly investigate leads. If X only preserves and retains this information for the minimum 90-day period, why does it do so when preserving this data longer could significantly enhance and prolong law enforcement's ability to investigate and prosecute child predators?

       **As described above, NCMEC, in its role as the primary interlocutor with law enforcement, is able to preserve the data provided by X as long as it deems necessary and in line with their retention policies. X has no objections to retaining the associated account data for longer than the current 90 day period should such an extension prove useful to law enforcement. This also highlights the importance of law enforcement having sufficient resources to investigate reports promptly.**

    c. Please confirm if X stores and retains the following information relating to reports to the CyberTipline:
        i. IP addresses **- YES**
        ii. Screen Names - **YES**

iii. User Profiles - **YES**
iv. Associated Screennames (by IP address and associated emails) - **YES**
v. Email addresses - **YES**
vi. Geolocation data - **YES**

d. If X does not retain or store any of the above types of information in question (c), please explain why.

**N/A**

e. Please list any other information X retains and preserves for law enforcement purposes not listed above in question (c).

**Any content and media associated with the CyberTipline report, including an entire archive of the user's account.**

f. Does X flag screennames and associated email addresses to suspected accounts that violate X's terms of service?

**Yes.**

2. How does X prioritize urgent requests for information from law enforcement and what is X's response time to urgent requests?

**X has a dedicated online portal for emergency information requests submitted by law enforcement. This portal is monitored 24/7 and all submissions are handled on a priority basis with responses in less than two hours.**

3. What is X's average response time to service of legal process from law enforcement for CSAM-related information?

**X endeavors to respond to legal process received from law enforcement and appropriate government entities in a prompt manner. Specifically, X endeavors to respond prior to the enumerated production date required by law in the particular jurisdiction or outlined in the individual legal process.**
**Upon identifying that a request seeks CSAM-related information, X prioritizes and handles such requests on an expedited basis.**

In 2023, the tech industry as a whole slashed more than 260,000 jobs. And in the first four weeks of this year, another 25,000 jobs were cut.

a. For each year, between 2018 and 2023, how many U.S. based employees did you have at X?

**2018 - Approximately 3920 employees globally**
**2019 - Approximately 4900 employees globally**
**2020 - Approximately 5500 employees globally**
**2021 - Approximately 7500 employees globally**
**2022 - Approximately 8000 employees globally (through Q3)**
**2023 - Approximately 1500 employees globally**

**We are continuing to investigate the breakdown of US based employees and will follow up with your staff.**

i. Of these employees, how many were sponsored on H-1B visas?

**In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

ii. For each year, between 2018 and 2023, how many H1-B visa applications did X submit?

**In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

b. For each year, between 2018 and 2023, how many employees based outside the U.S. did you have at X?

**In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

i. Of these employees, how many were based in China?

**The company does not have operations in China.**

c. For each year, between 2018 and 2023, how many employees in total did X terminate, fire, or lay off?
   i. Of these employees, how many were based in the United States?

**In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

ii. Did X fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?

**In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

iii. Were any duties and/or functions previously performed by laid-off employees transferred to or performed at any point by employees

sponsored on H1-B visas? If so, which duties and/or functions?

**In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

d. For each year, between 2018 and 2023, how many employees performing work related to child safety did X terminate, fire, or lay off?
    i. Of these employees, how many were based in the United States?

    **In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

    ii. Did X fill these newly vacant positions with employees sponsored on H1-B visas? If so, how many?

    **In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

    iii. Were any duties and/or functions (specifically relating to child safety) previously performed by laid-off employees transferred to or performed at any point by employees sponsored on H1-B visas? If so, which duties and/or functions?

    **In response to your inquiry, we are investigating this question and will be happy to follow up with your staff.**

    iv. How have layoffs impacted X's ability to protect children on its platforms?

    **As evidenced by the data on our enforcement, increased manual and automated reporting to NCMEC, implementation of advanced technologies, ten-fold increase in training of agents, and speed of execution on our number one priority, our capacity to combat CSAM has not been impacted, rather enhanced.**

    v. Does X have any plans to increase staff responsible for child safety operations or otherwise optimize its child safety operations?

    **We are building a Trust & Safety Center of Excellence in Austin, Texas, with a goal of hiring 100 full-time agents that will contribute to our safety operations.**

4. On January 30, 2024, the Tech Transparency Project (TTP) published an article on their website called, "Meta Approves Harmful Teen Ads with Images from its Own AI Tool". In summary, TTP, using Meta's "Imagine with Meta AI" tool generated inappropriate images such as young people at a pill party or other vaping. These images with text were

submitted to Facebook as advertisements targeting users between ages 13-17 in the United States. TTP reported that Facebook approved the advertisement, despite it violating its own policies, in less than five minutes to run on the following platforms: Facebook, Instagram, Messenger, and Meta Quest. Meta. Over the course of a week, TTP submitted the advertisements with the same end result: Facebook approving them. TTP reported that they canceled these advertisements before their scheduled publication, but it illustrated the repeated failures of Facebook to properly moderate content. This is just one example of what other non-government organizations and others have uncovered across social media platforms.

a. How often a month do X employees conduct quality checks on X's policies and safeguards for child accounts?

   **We are constantly evaluating our systems and protocols for ensuring the safety of our product. We do not allow minors under 13 to have accounts.**

b. In which departments, components, or units of the company does X have staff dedicated to performing this type of work?

   **Quality assurance and integrity of our work is built into our operational ethos and implemented across the company.**

c. How many employees make up these departments, components, or units?

   **Safety of minors is a shared responsibility across the entire company.**

d. If a violation is found, what action is taken, and how quickly is action taken?

5. Social media companies claim they are investing in company components dedicated to safety, and that their platforms are safe for children. However, children continue to be exploited daily across these platforms.
   a. What have X's revenue and profit figures been for the last three years (2021-2023)? Please provide figures broken out per year. Do not provide percentages.

      **Revenue for the years 2021 and first two quarters of 2022 were disclosed as part of Twitter's public company filings.**

      **2021 revenue was $5.08 billion, with a 2021 operating loss of $493 million and net loss of $221 million.**

      **2022 Q1 revenue was $1.2 billion, an operating loss of $128 million, and a net income of $513 million which includes a pre-tax gain of $970 million from the sale of MoPub for $1.05 billion and income taxes related to the gain of $331 million.**

**2022 Q2 revenue was $1.18 billion, with an operating loss of $344 million, and a net loss of $270 million.**

**Beginning in Q3 2022, X Corp., a private company, took over operation of the platform. As a privately-held company, X does not maintain or release public financial statements.**

b. How much has X spent in advertising for the last three years (2021-2023), broken out per year?

**As a public company in 2021, Twitter disclosed its annual financial statements as part of its public filings to the SEC. This mandatory disclosure included expenditures related to 'Sales and marketing'. Sales and marketing expenses consist primarily of personnel-related costs, including salaries, commissions, benefits and stock-based compensation for our employees engaged in sales, sales support, business development and media, marketing, corporate communications and customer service functions. In addition, marketing and sales-related expenses also include advertising costs, market research, trade shows, branding, marketing, public relations costs, amortization of acquired intangible assets, allocated facilities costs, and other supporting overhead costs. In its last annual disclosure, sales and marketing expenditures totaled $1,175,970,000.**

**Beginning in Q3 2022, X Corp., a private company, took over operation of the platform. X Corp. is a privately-held company and its budget allocations are confidential and competitively sensitive information.**

c. How much of X's resources spent on advertising has been devoted to advertising X's safety initiatives and efforts for the last three years (2021-2023), broken out per year?

**X does not have a specific budget line item for advertising safety initiatives.**

d. To get an understanding of how your company has invested and plans to invest in its components dedicated to child safety functions, what are the annual budgets for X's child safety-related components for the last three years (2021-2023)?

**X does not have a specific budget line item for 'child safety-related components,' as child safety is not limited to specific components and is prioritized across the entire Trust & Safety organization.**

e. What is the current anticipated (2024) budget for X's child safety-related components?

**X does not have a specific budget line item for 'child safety-related components,' as child safety is not limited to specific components and is**

**prioritized across the entire Trust & Safety organization.**

f. Provide the number of staff employed in X's child safety-related components for the last three years (2021-2023).

**X does not allocate a specific number of staff for child safety-related components.**

g. How much is that compared to X's other components for the same period? (Please provide a breakout per year. Do not provide percentages.)

**Our Trust & Safety teams are cross-functional and work across a variety of issues.**

h. How many staff are currently employed in X's child safety-related components?

**X does not allocate a specific number of staff for child safety-related components.**

i. What are the roles, responsibilities, and functions of X's child safety-related components?

**X has a combination of program managers, policy specialists, operations specialists, engineers, legal professionals, government affairs professionals, and other functions that work on issues related to child safety.**

j. Are any other components responsible for the monitoring of CSAM on X's platform(s)?

**See answer (i).**

k. What, if any, third parties does X employ or contract with to address CSAM material on its platforms?

**X partners with a network of vendors that supplement our Trust & Safety agent capacity.**

i. What are the roles and responsibilities of these third parties?

**Generally, the external Trust & Safety agent workforce conducts reviews of user reports, escalated and surfaced content, enforce our platform policies, and resolve account issues. This agent workforce undergoes regular and intensive training on our policies and enforcement guidance.**

ii. What is the breakdown of cost per third party over the last three years (2021-2023)?

**At this time we are not able to provide a specific cost breakdown of our partnerships with the network of Trust & Safety vendors.**

6. Of all reports sent by X to the National Center for Missing and Exploited Children, how many reports were self-generated from victim users for the last three years (2021-2023)? Please provide the actual number of self-generated reports in addition to the total number of reports (including those that were not self-generated). In addition, please provide a break-down of the self-reporters by age.

   **X does not capture whether the report is self-generated from victim users.**

7. What is X's policy or protocol with respect to law enforcement accessing user data and subsequent notification to users of law enforcement accessing their data?

   **Non-public information about X users will not be released to law enforcement except in response to appropriate legal process such as a subpoena, court order, other valid legal process, or in response to a valid emergency request.**

   **Requests for the contents of communications (e.g., posts or photos) require a valid search warrant or equivalent from an agency with proper jurisdiction over X.**

   **For purposes of transparency and due process, X's policy is to notify users (e.g., prior to disclosure of account information) of requests for their X account information, including a copy of the request, unless we are prohibited from doing so (e.g., an order under 18 U.S.C. § 2705(b)) or an exception applies (e.g., imminent threat to life, child sexual exploitation, or terrorism). We may ask that any non-disclosure provisions include a specified duration (e.g., 90 days) during which X is prohibited from notifying the user or may object to the non-disclosure order on free speech or other grounds.**

   a. Do certain crimes such as drug trafficking or child exploitation affect X's decision to notify a user whose data is accessed by law enforcement?

      **Yes. Exceptions to our user notice policy may include exigent circumstances or circumstances where our commitment to user notice is outweighed by the negative effects such notice would have on law enforcement efforts, such as emergencies regarding imminent threat to life, child sexual exploitation, or terrorism.**

   b. Do certain requests such as a subpoena or search warrant affect X's notification protocol? If so, what are they?

      **No, user notification rests on whether there is a non-disclosure provision in**

**the legal process provided by law enforcement and the presence of a policy exception.**

c. If X does notify users of law enforcement accessing their data, why does X find this necessary?

**For purposes of transparency and due process, X's policy is generally to notify users of requests for their X account information prior to disclosure of said account information.**

8. The National Center for Missing and Exploited Children has indicated that reports from social media companies tend to lack actionable information causing law enforcement to be burdened with incomplete information. How comprehensive are X's reports to NCMEC? What challenges is X experiencing on the collection of user data and other information to include in its reports to NCMEC? What actions is X taking to make its reports more comprehensive and useful to law enforcement?

**When we report a user to NCMEC, we include a full archive of that user's data for law enforcement to access via a subpoena for their investigation. From what NCMEC has shared with us, we provide significantly more user data than most companies, which helps law enforcement prosecute violators. We are working on a variety of changes to our system to increase data throughput and reduce NCMEC submission error rates.**

**Senator Mike Lee**
**Questions for the Record**
**Linda Yaccarino, Chief Executive Officer, X Corp.**
**Hearing on "Big Tech and the Online Child Exploitation Crisis"**
Submitted February 7, 2024

1. In last week's hearing, you stated that "[i]t's time to criminalize the sharing of nonconsensual material." Will you support the PROTECT Act of 2024, S. 3718, which creates a pathway for victims of nonconsensual image dissemination to easily have their images removed and creates civil and criminal liability for platforms who knowingly share those nonconsensual images?

   **We support the SHIELD Act, and we will evaluate the PROTECT Act and look forward to discussing it with you and your staff.**

2. The 2022 Thorn Report indicated that 19 percent of minors who use X have had an online sexual interaction on your platform, and 13 percent of minors who use X had an online sexual interaction with someone they believed to be an adult. What are you doing to put an end to these interactions?

   **In early 2023, after the acquisition, we implemented safety-by-default features for accounts opened by users between the ages of 13-17. Accounts belonging to known minors are now defaulted to a "protected" setting. This means that known minors will receive a request when new people want to follow them (which they can approve or deny), that their posts will only be visible to their followers, and that their posts will only be searchable by them and their followers (i.e. they will not appear in public searches). Under this setting, accounts belonging to known minors will be restricted to receiving DMs from accounts they follow by default. We also utilize an age lock. Once a new user enters a date of birth that makes them under the age of 18, they will be stopped from re-entering a new date of birth for that account.**

   **We also take steps to limit exposure to sensitive content. Known minors or viewers who do not include a birth date on their profile are restricted from viewing specific forms of sensitive media such as adult content. X obscures sensitive media behind notices and interstitials. This includes our product age restrictions that restricts known minors from viewing adult content.**

   **In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.**

**In addition, X automatically excludes potentially sensitive media (along with accounts users have muted or blocked) from search results shown to accounts of known minors or without a date of birth.**

**More information on our protected account settings can be found in our Help Center.**
**https://help.twitter.com/en/safety-and-security/public-and-protected-posts**

3. X's policies permit pornography on your platform. An estimated 13 percent of total material on X is explicit. How do you guarantee that minors cannot see any of this material on your platform?

   **X utilizes content filters for sensitive media and minor account holders are restricted from seeing content marked as sensitive. X may also use automated techniques to detect and label potentially sensitive media, and to detect and label accounts that frequently post potentially sensitive media.**

4. X restricts certain content from accounts that belong to minors. However, the only age verification measure that X undertakes to ascertain the age of its users is asking new users to enter their birthdate when they open an account. How do you prevent minors from lying about their age when creating an account?

   **X's age assurance process relies on self-declaration to collect the user's date of birth through the neutral presentation of a date of birth prompt and allows any user to report other users who they believe are under the age of 13. X has set up a dedicated age moderation workflow to enable any user to report an account that they suspect is being used by a minor under the age of 13, available at https://help.twitter.com/en/forms/safety-and-sensitive-content/underage-user.**

   **X strongly supports app stores handling age-gating for apps. Age verification through app stores is simply leveraging existing processes, enhances teen safety by filtering all inappropriate apps and preserves privacy by avoiding the need for personal information sharing at the individual app level.**

5. Currently, X only encrypts private messages for users with premium accounts. If a minor upgrades to a premium account, will you encrypt their messages as well? Will this hamper your ability to detect sexual exploitation of minors or grooming behaviors?

**Premium subscriptions require credit card information, which minors should not have unless authorized by a parent. Direct Messages are not encrypted by default, rather premium subscribers must opt into a new encrypted message, and only text is encrypted, which allows us to continue scanning DM surfaces for violative content and CSAM.**

6. In 2021, there was a widely-reported failure by Twitter to suspend accounts sharing CSAM and remove nonconsensual videos of a 13-year-old minor that were shared on your platform. The minor had been a victim of sextortion, and after he had ceased communicating with the predators, they shared his images on Twitter. The minor became aware of his images which appeared on two well-known CSAM Twitter accounts and had been viewed nearly 200,000 times. The minor made three separate requests to Twitter to have his images removed—including one in which he sent a copy of his ID to verify that it was him in the videos. Despite these efforts, the minor was told that in Twitter's review of those videos and those accounts, you "didn't find a violation of our policies." Twitter only acted on those videos and accounts after a federal agent made a "take-down demand." This event happened prior to Twitter's acquisition by Elon Musk and the shift in Twitter's, now X's, priorities. However, in May of 2023, researchers at the Stanford Internet Observatory marked 40 images of CSAM that remained on X's platform over a period of two months, despite those images being marked at CSAM by Microsoft's PhotoDNA tool and existing in NCMEC's photo-hashing database. What is the current process for a person to have their nonconsensual images removed from X? What is the maximum amount of time you permit those images to remain before removal? What changes have you made since 2021 to ensure that the 13-year-old minor's experience will never be repeated?

   **X maintains a policy against the sharing of non-consensual intimate media.**

   **You may not post or share intimate photos or videos of someone that were produced or distributed without their consent.**

   **Sharing explicit sexual images or videos of someone online without their consent is a severe violation of their privacy and the X Rules. Sometimes referred to as revenge porn, this content poses serious safety and security risks for people affected and can lead to physical, emotional, and financial hardship.**

   **Under this policy, you can't post or share explicit images or videos**

**that were taken, appear to have been taken or that were shared without the consent of the people involved.**

**Examples of the types of content that violate this policy include, but are not limited to:**

- **hidden camera content featuring nudity, partial nudity, and/or sexual acts;**
- **creepshots or upskirts - images or videos taken of people's buttocks, up an individual's skirt/dress or other clothes that allows people to see the person's genitals, buttocks, or breasts;**
- **images or videos that superimpose or otherwise digitally manipulate an individual's face onto another person's nude body;**
- **images or videos that are taken in an intimate setting and not intended for public distribution; and**
- **offering a bounty or financial reward in exchange for intimate images or videos.**

**We will immediately and permanently suspend any account that we identify as the original poster of intimate media that was created or shared without consent. We will do the same with any account that posts only this type of content, e.g., accounts dedicated to sharing upskirt images.**

**In other cases, we may not suspend an account immediately. This is because some people share this content inadvertently, to express shock, disbelief or to denounce this practice. In these cases, we will require you to remove this content. We will also temporarily lock you out of your account before you can post again. If you violate this policy again after your first warning, your account will be permanently suspended.**

**We are also evaluating the technical requirements to participate in NCMEC's Take It Down program.**

7. What is X doing to prohibit known CSAM from being uploaded and shared on your platform?

**Media hash matching is one of the ways we take down instances of known CSAM circulating on the platform. In December 2022, we launched a new hash matching pipeline called Safer through our partnership with Thorn that allows us to take down more media than before. Through Thorn we also have access to a CSAM media classifier for the first time, which allows us to detect previously unseen images. In 2023, we detected over 60,000 pieces of media through Safer.**

**We also leverage hash databases that are maintained by NCMEC and the Tech Coalition to detect known CSAM on X. We scan all media uploaded to X for matches to hashes of known CSAM sourced from NGOs, law enforcement and other platforms. Users posting known content are immediately permanently suspended and reported to NCMEC.**

8. Predators often engage with a minor and then quickly attempt to move the conversation to another app with more encryption protections. What does X do to identify these types of interactions, and what does X do to prevent them? Do you report to NCMEC when you suspect that a predator is engaged in grooming a minor?

   **Yes, we do report this behavior to NCMEC. Also, we recently submitted our application to join Project Lantern, a program developed by the Tech Coalition. The purpose of this initiative is to enable tech and related industry companies to share information and signals to root out cross-platform bad actors.**

9. Do you provide information provided to NCMEC regarding suspected grooming or sexual abuse to a minor's parents?

   **No, as we do not collect information of parents as they are not linked to accounts of minors.**

**SENATOR TED CRUZ**

**U.S. Senate Committee on the Judiciary**

**Questions for the Record for Linda Yaccarino, CEO, X**

## I. Directions

Please provide a wholly contained answer to each question. A question's answer should not cross-reference answers provided in other questions.

If a question asks for a yes or no answer, please provide a yes or no answer first and then provide subsequent explanation. If the answer to a yes or no question is sometimes yes and sometimes no, please state such first and then describe the circumstances giving rise to each answer.

If a question asks for a choice between two options, please begin by stating which option applies, or both, or neither, followed by any subsequent explanation.

If you disagree with the premise of a question, please answer the question as-written and then articulate both the premise about which you disagree and the basis for that disagreement.

If you lack a basis for knowing the answer to a question, please first describe what efforts you have taken to ascertain an answer to the question and then provide your tentative answer as a consequence of its reasonable investigation. If even a tentative answer is impossible at this time, please state why such an answer is impossible and what efforts you intend to take to provide an answer in the future. Please further give an estimate as to when Senator Cruz will receive that answer.

To the extent that an answer depends on an ambiguity in the question asked, please state the ambiguity you perceive in the question, and provide multiple answers which articulate each possible reasonable interpretation of the question in light of the ambiguity.

## II. Questions

1. In the last two years, has an employee or commissioner of the Federal Trade Commission (FTC) requested to evaluate or evaluated your data used for training Large Language Models or algorithms or the sources of such data for bias, discrimination, or misinformation?

   **No.**

2. In the last two years, has an employee or commissioner of the FTC sought details regarding your company's measures related to filtering or blocking inputs and outputs of a Large Language Model or algorithms.

   a. If yes, has the FTC attempted to coerce or otherwise request you to implement input/output filtering in order to allegedly comply with federal law?

   **No.**

3. In the last two years, has an employee or commissioner of the Federal Trade Commission sought to evaluate your company's use of measures, including "prebunking" or "debunking", designed to counteract so called "online misinformation"?

   **No.**

4. In June 2022, the FTC released a report titled "Combatting Online Harms Through Innovation." In this report, the FTC discussed how the deployment of AI tools intended to detect or otherwise address harmful online content is accelerating but may never be appropriate as an alternative to human judgment.

   a. In the context of protecting children from online harms to what extent does your company rely on automated tools to detect online harm vs. human review? Please be specific.

      **In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year. For the first time ever, we are evaluating all videos and GIFs posted on X for CSAM. Since**

**launching this new approach in July 2023, we have matched over 70,000 pieces of media.**

**Our proactive detection efforts are primarily driven by the following tools:**

- **Hash-sharing: Our current methods of surfacing potentially violating content for human review include leveraging the hashes provided by NCMEC and industry partners - which makes this the most widely used form of CSAM detection. We scan all media uploaded to X for matches to hashes of known CSAM sourced from NGOs, law enforcement and other platforms. Users posting known content are immediately permanently suspended and reported to NCMEC. For videos we use a proprietary hashing algorithm produced by Thorn.**
- **Automatic text detection: We have a variety of tools to access the likelihood that a post is advertising or promoting the sharing of child sexual abuse material. Some of these defenses lead to automatic suspensions while other users are flagged for human review.**
- **For videos we use a proprietary hashing algorithm produced by Thorn.**
- **PhotoDNA and internal proprietary tools: a combination of technology solutions are used to surface accounts violating our rules on Child Sexual Exploitation.**
- **Media Risk Scanning: We receive a media classifier score through Safer and it is used to filter false positive hash matches at the moment. We use a novel classifier model to rate media shared through posts' likelihood of being CSAM. Media that receives a high score is then sent to human review.**

  **In addition, X is currently beta testing a text-based machine learning classifier developed by Thorn to assist in the detection of sextortion, CSAM, child-access, and child sexual abuse discussion.**

b. What benefits can AI provide to helping detect and/or stop harmful content to children online?

**AI can assist in proactive detection, hash-matching, analysis of**

**media, analysis of text, and automation of enforcement, to name a few benefits.**

c. What does a human reviewer provide that an AI or automated tool cannot? Will we always need some measure of human review in assessing online harms to children?

**Human capacity is essential for understanding the context of cybercrimes, investigating account behaviors, analyzing networks of bad actors, providing qualitative feedback on processes, and working with cross-functional teams to enforce policies.**

d. The FTC has sent mixed signals in its enforcement of COPPA. While the Commission emphasizes not over relying on use of automated tools or AI, they have nonetheless found liability for using human review as alternative signaling overreliance on automated tools. What improvements, if any, should Congress make to clarify the legal tension between use of automated detection tools vs. human review?

**X has billions of accounts on its platform; robust, automated tools are essential to implement its safety policies—including policies related to COPPA—at scale. We encourage regulators and lawmakers to create the space for platforms such as X to innovate and improve these automated tools, including by protecting good faith efforts to improve online safety from legal challenges.**

5. In 2021, Congress directed the FTC to research and report on how AI can be used positively to detect and combat fraudulent or deceptive content online. Rather than viewing AI as a potential solution to our online woes, the FTC instead issued a report that read more like an indictment of the technology.

   a. Please explain whether, in your view, AI can be used to positively detect and combat fraudulent or deceptive content, including the recent use of deepfakes or other scams to harm consumers.

   **Yes.**

   b. Has the FTC ever consulted with your company to learn how your company deploys AI to better detect and combat fraudulent or deceptive content? Has the DOJ? How about the Federal Elections Commission?

**To the best of our knowledge, X has not been consulted on these issues.**

c. How can Congress empower agencies to use AI positively for the protection of American consumers from fraudulent or deceptive content?

**AI is an important and developing technology for detecting and combatting fraudulent and deceptive content. Both the private sector and government agencies should carefully assess how they can deploy that evolving technology most effectively. Congress can help to ensure that federal agencies have the personnel and resources necessary to evaluate and deploy that technology most effectively.**

6. Please provide a description of your company's policy regarding the sale or transfer of the data of American users collected on your platform to a third party, including data brokers.

**As stipulated in Section 6 of our Privacy Policy (https://twitter.com/en/privacy), we have specific legal bases for collecting, using and sharing user data (further described here https://help.twitter.com/en/rules-and-policies/data-processing-legal-bases) but do not sell user's personal data.**

7. Has your company ever sold the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the categories of data sold.

**In response to your inquiry, our teams are investigating this question and will be happy to follow up with you.**

8. Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to the government of a foreign country? If so, please provide a full list of the countries and the circumstances underlying the basis for such transfer.

**Not to our knowledge, with the exception of an incident several years ago under prior management in which malicious actors inside the company transferred data concerning a discrete number of users to the government of Saudi Arabia without the company's knowledge or consent. The company cooperated with U.S. authorities in connection with the incident.**

9. Has your company ever sold the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and the categories of data sold.

**In response to your inquiry, our teams are investigating this question and will be happy to follow up with you.**

10. Outside of complying with a lawful order, has your company ever transferred the data of American users on your platform to a U.S. government agency? If so, please provide a full list of the agencies and categories of data transferred.

**Not to our knowledge.**

11. Does your company have a policy to restrict third party use and/or transfer of data collected from users on your platform? Please be specific, including how you enforce such restrictions and whether such restrictions prohibit the sale or transfer of such data to a government agency, including a foreign government agency.

**X only permits third-party access to aggregate data via our approved API process. When a developer seeks access to X's API, they are required to abide by the terms of our developer agreement upon sign up. (ttps://developer.twitter.com/en/developer-terms/agreement-and-policy), a legally binding agreement between the developer and X. X specifically prohibits the use of user data by any entity for surveillance purposes, investigating or tracking X users or their content or in any other way inconsistent with users' privacy expectations as described in our Privacy Policy (https://twitter.com/en/privacy). Should we learn that a developer has violated these policies, we will take appropriate action, which may include suspension and termination of access to X's APIs, reporting to relevant authorities, and/or legal action as appropriate. Further information on restricted uses of X's API can be found at https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases.**

12. Between July 4, 2023 and July 14, 2023, was your company contacted by any employee of or contractor for any of the following agencies? Please answer "yes" or "no" for each agency and, if "yes," provide the date(s) of contact and the name(s) of the agency employees or contractors that contacted your company.

**X's responses to questions 12(a)-(l) are based on a reasonable search of its systems and exclude routine requests for support with X accounts.**

a. U.S. Department of Health and Human Services (HHS)

**X is not aware of any outreach by an employee or contractor of HHS during the specified time period.**

b. National Institute of Allergy and Infectious Diseases (NIAID)

**X is not aware of any outreach by an employee or contractor of NIAID during the specified time period.**

c. Centers for Disease Control and Prevention (CDC)

**X is not aware of any outreach by an employee or contractor of CDC during the specified time period.**

d. U.S. Food and Drug Administration (FDA)

**X is not aware of any outreach by an employee or contractor of FDA during the specified time period.**

e. The National Institutes of Health (NIH)

**X is not aware of any outreach by an employee or contractor of NIH during the specified time period.**

f. U.S. Department of Homeland Security (DHS)

**X is not aware of any outreach by an employee or contractor of the DHS during the specified time period.**

g. DHS Cybersecurity and Infrastructure Security Agency (CISA)

**Certain X employees received mass emails sent to CISA listservs, including notifications of trainings and public cybersecurity vulnerability notifications. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of CISA during the specified time period**

h. U.S. Census Bureau

**X is not aware of any outreach by an employee or contractor of the Census Bureau during the specified time period.**

i. Federal Bureau of Investigation (FBI)

**X engaged in communications with the FBI related to routine law enforcement requests, and received notifications of webinars and other industry-wide training opportunities, during the specified time period. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of the FBI during the specified time period.**

j. U.S. Department of Justice (DOJ)

**X engaged in communications with the DOJ related to routine law enforcement requests during the specified time period. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of the DOJ during the specified time period.**

k. The White House Executive Office of the President (EOP)

**X is not aware of any outreach by an employee or contractor of the EOP during the specified time period.**

l. U.S. Department of State

**Certain X employees received emails from State Department listservs during the specified time period. None of these communications related to X's content moderation practices, policies, and/or procedures. Apart from these communications, X is not aware of any outreach by an employee or contractor of the Department of State during the specified time period**

13. Is it your company's policy to prevent children under 13 from using your social media app(s) or creating an account?

**Yes.**

14. In your view, would it be appropriate for school-aged children to spend time on or access your company's social media app(s) during class?

**Technology usage in the classroom is a matter of concern for the school district, school administration, the teacher, and parents.**

15. As a parent, would you be concerned if your child were able to access your company's social media app(s) during class via a school network or device?

**Technology usage in schools should be managed by school administration, teachers, and parents.**

16. In your view, should elementary and secondary schools block students' access to your company's social media app(s) on school networks and devices?

**We defer to school administrators on which technologies to allow and which apps to block on their respective networks and devices.**

17. Do you think that school buses equipped with Wi-Fi should allow children to access your company's social media app(s) via a school bus Wi-Fi network during their rides to and from school?

**We defer to school administrators on which technologies to allow and which apps to block on their respective networks and devices.**

18. As a parent, do you think it is important to supervise your children's internet access?

**Yes.**

19. As a parent, would you be concerned if your child's school allowed your child to access the internet on an unsupervised basis, such as on your child's bus ride to and from school via the school bus Wi-Fi?

**Yes.**

20. Do you think Congress should require schools, as a condition of receiving broadband subsidies through the Federal Communications Commission's E-Rate program (which funds broadband for elementary and secondary schools), to block students' access to your company's social media app(s) from school-run networks?

**We generally do not oppose efforts to regulate access to social media platforms in schools or on school-run networks.**

21. Do you support the bipartisan *Eyes on the Board Act of 2023*, S. 3074?

**While we have not taken an official position on this legislation, we generally have not opposed legislative efforts regulating access to social media platforms in schools.**

22. Have you, your company, or any foundation associated with you or your company, donated or contributed funding, equipment, or services to any of the following organizations in the last ten years (CY 2013 to CY 2023)?

    a. Education and Libraries Networks Coalition (EdLiNC) - **No**

    b. Open Technology Institute - **No**

    c. Consortium for School Networking (COSN) - **No**

    d. Funds For Learning - **No**

    e. State Educational Technology Directors Association (SETDA) - **No**

    f. Schools, Health, and Libraries Broadband Coalition (SHLB) - **No**

    g. State E-Rate Coordinators' Alliance (SECA) - **No**

    h. EducationSuperHighway - **No**

    i. All4Ed - **No**

    j. Public Knowledge - **Yes**

    k. Fight for the Future - **No**

    l. Free Press - **No**

    m. Electronic Frontier Foundation - **No**

    n. Benton Foundation or Benton Institute for Broadband & Society - **No**

    o. Electronic Privacy Information Center - **No**

23. For each such donation or contribution described in the prior question, please detail (1) the type of donation or contribution, such as financial donation, goods or equipment, services, etc.; (2) who made the donation or contribution; (3) the recipient organization; (4) the year the donation or contribution was made; and (5) the total value of that donation or contribution.

**Public Knowledge - $10,000 in 2021 donated from Twitter**

**24.** Yes or no: Did employees of or contractors for the Cybersecurity and Infrastructure Security Agency (CISA) ever ask Twitter/X to meet with employees of or contractors for the Department of Homeland Security Office of Inspector General (DHS OIG)?

**To the best of my knowledge, X has not received any such request from CISA.**

**a.** If yes, provide the date of the request from CISA, the channel through which the request was made, and the name of the CISA employee(s) or contractor(s) who made the request.

**Questions from Senator Thom Tillis**
**for the CEO of X Corp., Ms. Linda Yaccarino**

1. Twenty-one is the minimum age to purchase highly regulated adult products such as alcohol, tobacco, and nicotine. Nevertheless, there is a proliferation of user-generated content posted on social media sites featuring underage use of these products.

   Recently, some have proposed banning these age-restricted products due in part to the user-generated content being available on your respective platforms. Surely, banning these products cannot be the answer. However, we must do more – your company must do more – to shield underage audiences from exposure to this content.

   Therefore, as the content moderator of these platforms, what policies do you have in place, and what more can you do, to prevent this type of user-generated content from reaching underage audiences? How do you respond to requests to pull this content from your sites when deemed inappropriate for underage audiences?

   **X prohibits knowingly marketing or advertising the following products and services to minors:**

   - **Alcoholic beverages and related accessories**
   - **Weapons, ammunition, or weapons training/certification**
   - **Projectile, BB, or pellet guns/devices***
   - **Fireworks***
   - **Aerosol paint or etching cream capable of defacing property**
   - **Tobacco products or accessories, including electronic cigarettes***
   - **Any controlled substance or paraphernalia***
   - **Drug paraphernalia***
   - **Any substance or material containing Salvia divinorum or Salvinorin A***
   - **Weight loss products and services and content focused on weight loss**
   - **Health and wellness supplements (including, but not limited to, health, dietary, food, nutrition, weight loss, and muscle enhancement substances and supplements)**
   - **Tanning in an ultraviolet tanning device**
   - **Gambling products and services, including lotteries**
   - **Body branding such as tattooing, body piercing, or permanent cosmetics**
   - **Sexual products and services, or content that is adult in nature**
   - **Please note that asterisked items are prohibited from advertising on X overall.**

   **While posts promoted through X's advertising services are labeled as "Promoted" and must abide by our X Ads Policies, organic, non-promoted posts may also be considered paid product placements, endorsements, or advertisements ("Paid Partnerships").**

**The following are examples of Paid Partnerships:**

- **A user, including a creator or brand, has been or may be compensated for a post (including in the form of money, gifts, loans of products, or other rewards or incentives), or**
- **A post is created as part of, or in connection with, a commercial relationship (such as a current or recent 'brand ambassador' arrangement), or**
- **A post includes an affiliate link or discount code through which the user might receive some kind of benefit, incentive or reward**

**Posts that are part of a Paid Partnership posted as an organic post will require clear and prominent disclosures indicating the commercial nature of such content. For example, "#ad", "#paidpartnership", "#sponsored".**

**Failure to include an appropriate disclosure in a clear and prominent way could result in enforcement actions. In addition to abiding by the X Rules, users, including creators and brands, that participate in Paid Partnerships are responsible for complying with all applicable laws and regulations, including but not limited to, all advertising laws and, where applicable, FTC regulations including the FTC's Guides Concerning the Use of Endorsements and Testimonials in Advertising.**

2. Public reports conclude that drug cartels use social media like TikTok, META, X, Snapchat, and others to plan, organize, and communicate in real-time. These communications coincide directly with criminal activity.

   What are your companies doing to crack down on cartel coordination? Specifically, in the recruitment of children to commit crimes or assist in the sale/distribution of illicit drugs?

   **We are aware of general reports of cartels utilizing certain social media platforms to advertise to and recruit minors, however, we are unaware of any such coordinated campaigns on X. We remain vigilant of this activity and if your office or any stakeholder has evidence of this behavior on X, please share immediately with our teams so that we may investigate.**

   **X maintains a robust cybercrime policy on Illegal and Regulated Goods (IRGS) and Services. Under this policy, users may not use our service for any unlawful purpose or in furtherance of illegal activities. This includes selling, buying, or facilitating transactions in illegal goods or services, as well as certain types of regulated goods or services.**

   **In addition to reports received, we proactively surface activity that may violate this policy for human review.**

**Goods or services covered under this policy include, but are not limited to:**

- **counterfeit goods and services;**
- **drugs and controlled substances;**
- **human trafficking;**
- **products made from endangered or protected species;**
- **sexual services;**
- **illicitly obtained materials; and**
- **weapons, including firearms, ammunition, and explosives, and instructions on making weapons (e.g. bombs, 3D printed guns, etc.**

**If we determine that a user violated this policy, we may suspend their account, including upon first review.**

**Accounts that appear to be using misleading account information in order to engage in spamming, abusive, or disruptive behavior to promote the sale of illegal and regulated goods and/or services may be subject to suspension under our platform manipulation and spam policy.**

3. What steps does your platform take to proactively remove, delist, and ban any posts, users, websites, and advertisements associated with the sale and distribution of fentanyl and other illicit drugs?

**Under our IRGS policy outlined above, we use a mix of proactive detection and user reporting to enforce potential violations. We proactively detect content for human review using a mix of technology and heuristics depending on the activity. We utilize databases of known terms and slang, drug names, and emojis, for example, to inform our detection models. If a particular link is found to be harmful, we denylist that URL and block it from being posted at all. More information on our approach to harmful links can be found here: https://help.twitter.com/en/safety-and-security/phishing-spam-and-malware-links**

4. One area of growing concern is the sale and distribution of fake or counterfeit vaping devices online, particularly in connection with so-called Delta-8 THC. Counterfeit vapes, many coming from China, have unsafe and even potentially deadly chemicals. They have caused hospitalizations and death. What are your platforms doing to combat this problem?

**See above answer to Question 1 regarding prohibited advertising to minors, which captures tobacco accessories, including e-cigarettes. Such activity could also be captured by our Illegal and Regulated Goods Policy, as outlined in our answer to Question 2 above.**

5. What are the main impediments your platform encounters in identifying all fentanyl and illicit drug advertisements posted to your platform(s) automatically? Please describe any circumstances in which you do not or cannot apply detection

technologies against content transmitted on your platform(s).

**We are constantly soliciting input from experts and law enforcement on the latest trends, nomenclature, and slang involving the trafficking of illegal goods and services. The information we receive from these entities feeds into our proactive detection capabilities.**

6. How many posts, users, websites, and advertisements have you removed, delisted, and banned per year for the sale and distribution of fentanyl and other illicit drugs? How many per year? Have you seen an increase in illicit drugs being advertised to children on your platform(s)?

**In 2023, we suspended approximately 630,000 accounts and removed more than 1.8 million posts under our Illegal and Regulated Goods Policy.**

7. Are there any other roadblocks or impediments that you face in addressing fentanyl and illicit drug advertisements on your platform(s), and working with law enforcement on such matters? If yes, what are they? If no, how many cases have been transmitted to law enforcement and DEA?

**We are constantly soliciting input from experts and law enforcement on the latest trends, nomenclature, and slang involving the trafficking of illegal goods and services. The information we receive from these entities feeds into our proactive detection capabilities.**

**In response to your inquiry regarding transmissions to law enforcement, our teams are investigating this question and will be happy to follow up with you.**

8. How do you work with organizations, advocates, and experts focused on drug prevention and addiction recovery to adapt your products and operations to keep up with the illicit drug crisis — including working with parents that have lost children due to lethal drugs bought online?

**We work with organizations around the world dedicated to supporting recovery and online safety. One notable partner in the US in this work is Mobilize Recovery, and we are honored to support their campaigns on X via advertising credits.**

9. What are the total number of meetings that your company has had with parents to address online safety concerns? Can you provide the total number of meetings over the last three years? Please separate this last question's answer by number per year.

**We have had numerous meetings around the globe with groups dedicated to the safety of minors online, including parents. We do not have a specific number of meetings we are able to share. We are committed to continuing to meet with**

**advocates for child protection and gather feedback on how we can make X safer for minors.**

10. In 2022, then National Center for Missing & Exploited Children (NCMEC) received over 32 million reports of Child Sexual Abuse Material (CSAM). Reports of online sex crimes to the CyberTipline are growing exponentially year by year. Out of those 32 million reports, how many did your platform submit to NCMEC?

**In 2022, Twitter submitted approximately 98,000 reports to the CyberTipline. In 2023, X submitted approximately 850,000 reports to the CyberTipline.**

**In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year.**

**Since April of 2023, we have increased training for content moderators on the tools and policies for NCMEC reporting. In turn, this has led to a 10x increase in the volume of manually-submitted NCMEC reports, from an average of 6,300 reports per month to an average of 64,000 reports per month from June through November 2023. We are evaluating more sources of potential CSAM than we could before.**

11. There is concern that this number is going to fall dramatically this year because of the adoption of end-to-end encryption, not because the problem is going away. How will your company track and address this issue moving forward?

**At this time, X does not offer full end-to-end encryption. Encrypted Direct Messages are limited to only subscribers of X Premium, and only text is encrypted. Encrypted conversations will appear as separate conversations, alongside your existing Direct Messages in your inbox. Direct Messages are not defaulted to encrypted.**

**Users need to satisfy the following conditions in order to send and receive encrypted messages:**

i. **both sender and recipient are on the latest X apps (iOS, Android, Web);**
ii. **both sender and recipient are verified users or affiliates to a verified organization; and**
iii. **the recipient follows the sender, or has sent a message to sender previously, or has accepted a Direct Message request from the sender before.**

12. Has your platform seen an increase of suspected online child sexual exploitation-CSAM over the past few years? If so, what do you believe is the driving factor on why it's happening on your platform?

**Generally, we have not seen a marked increase in the prevalence of CSAM, however, we have improved our capabilities of detecting CSAM, made it easier to report CSAM across all product surfaces, increased training of agents, implemented more automated technologies, and become more aggressive in automated enforcement of accounts that engage with the content.**

13. What are some new tools or strategies that your platform has implemented to identify CSAM? How closely does your platform work with NCMEC?

**We are constantly seeking feedback and input from trusted organizations that are aligned in the mission to combat online CSE. Foundational to our work is our multidimensional partnership with NCMEC, which manages the CyberTipline program, regularly convenes global stakeholders and facilitates actionable feedback from law enforcement that makes us better. We have quarterly operational and policy syncs and members of our teams are connecting every month to share information and feedback on our reporting. Other instrumental partners are the Tech Coalition and WeProtect, alliances that push our innovation and provide critical information sharing on emerging threats and behaviors.**

**In December 2022, we launched a new product partnership that allows us to take down more violative media than before. Built by Thorn, Safer allows tech platforms to identify, remove, and report child sexual abuse material at scale.**

**We are investing in products and people to bolster our ability to detect and action more content and accounts, and are actively evaluating advanced technologies from third-party developers that can enhance our capabilities. Some highlights include:**

a. **Automated NCMEC reporting: In February 2023, we sent our first ever fully-automated NCMEC CyberTipline report. Historically, every NCMEC report was manually reviewed and created by an agent. Through our media hash matching with Thorn, we now automatically suspend, deactivate, and report to NCMEC in minutes without human involvement. This has allowed us to submit over 50,000 automated NCMEC reports in the past year.**
b. **Expanded Hash Matching to Videos and GIFs: For the first time ever, we are evaluating all videos and GIFs posted on X for CSAM. Since launching this new approach in July 2023, we have matched over 70,000 pieces of media.**
c. **Launched Search Intervention for CSE Keywords: CSAM impressions occur more on search than on any other product surface. In December 2022, we launched the ability to entirely block search results for certain terms. We have since added more than 2,500 CSE keywords and phrases to this list to prevent users from searching for common CSE terms.**

14. What resources or help does your platform provide to victims of CSAM? Does your platform work with local victim groups and professionals?

**The uniqueness of X is the role it serves as a platform for public conversation, the global town square of the internet. X has always been a place for victims to bring awareness to their causes and issues of public concern, like legislation. We will continue to support organizations around the world that promote online safety and we welcome any recommendations of victims groups that we could support in their campaigns and advocacy.**

15. What are the top technical hurdles your company faces in combating CSAM?

**Cross-platform bad actors always pose a challenge as they may attempt to direct people to other platforms where CSAM is exchanged or transactions happen. Sharing signals between platforms will assist in rooting out these types of behaviors, and that is why we have applied to the Tech Coalition's Project Lantern.**

16. There seem to be competing views on how to regulate algorithms. Some suggest that more transparency is needed, while others want more privacy. Can you provide your perspective on whether more or less transparency is needed when it comes to algorithms?

**In 2023, we published our recommendation algorithm and a comprehensive blog explaining our recommendation system.**
**https://blog.twitter.com/en_us/topics/company/2023/a-new-era-of-transparency-for-twitter**

**You can find the code to our recommendation algorithm on GitHub.**
**https://github.com/twitter/the-algorithm**

17. Do you believe that large companies and platforms like yours can use algorithms to stifle innovation or small businesses?

**We have published our recommendation algorithm in an effort to be transparent about how our systems work and to encourage innovation.**

18. What do you believe is the role of government in regulating algorithms? What, if any, unintended consequences would there be if Congress gets involved?

**Foundational to regulation of algorithms, or artificial intelligence, is strong privacy protections for individuals. Congress should begin by passing comprehensive privacy reform. We also believe it is important to give people control, which is why we offer all users the 'Following' experience, as an alternative to the algorithmic 'For You' tab.**

19. Are you aware of your platform using surveillance advertisements to target children (anyone under the age of 18) with specific ads? If so, in your opinion, how can this be mitigated?

**To the best of my knowledge, X does not use surveillance advertisements to target children with specific ads. Generally, you may not specifically target advertisements to 13-17 year olds on X.**

20. Beyond surveillance advertisements, are there any other algorithmic-based practices being implemented that are particularly detrimental to children? In your opinion, how can this be mitigated?

**To the best of my knowledge, I am not familiar with algorithmic-based practices being implemented that are particularly detrimental to children.**

21. Are you aware of any surveillance advertisements or algorithms that are used to target children, specifically to promote drugs and the sale of narcotics?

**To the best of my knowledge, I am not aware of surveillance advertisements or algorithms that are used to target children on X.**