

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mr. Dario Amodei
Chief Executive Officer
Anthropic
548 Market Street
San Francisco, CA 94104

Dear Mr. Amodei:

We write today to request information on Anthropic's efforts to implement sufficient protections to prevent scammers from misusing its technology. In recent years, global criminal networks have turned to AI to target more people with increasingly personalized and believable digital scams, contributing to a booming scam economy that is now a larger illicit industry than the global drug trade.¹ The FBI recorded \$16.6 billion in reported losses to suspected scams and other cybercrime in 2024 – up from \$3.5 billion in 2019, a more than 370 percent increase over the last five years.² The federal government has a responsibility to protect the American people from scams, but this effort requires an all-hands-on-deck approach across multiple industries. We hope we can work together productively on this important issue.

With advancements in AI, scams will continue to grow in sophistication, frequency, and impact.³ In the early phases of a scam, criminals can use generative AI to quickly identify and then collect details on their targets, enabling them to create tailor-made scams.⁴ Once armed with

¹ *Online Scams May Already Be as Big a Scourge as Illegal Drugs*, The Economist (Feb. 6, 2025) (www.economist.com/briefing/2025/02/06/online-scams-may-already-be-as-big-a-scourge-as-illegal-drugs); The Economist Group, *Press Release: Fraud Factories, Cyber Criminals and Corruption: The Economist's New Podcast, "Scam Inc", Uncovers a New, Global, Underground Economy Worth More Than Illicit Drug Trade* (Feb. 6, 2025) (www.economistgroup.com/press-centre/the-economist/fraud-factories-cyber-criminals-and-corruption-the-economists-new-podcast-scam-inc).

² FBI, *Internet Crime Report* (2024) (www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf); FBI, *Internet Crime Report* (2019) (www.ic3.gov/AnnualReport/Reports/2019_ic3Report.pdf).

³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

⁴ Fredrik Heiding, Bruce Schneier, and Arun Vishwanath, *AI Will Increase the Quantity — and Quality — of Phishing Scams*, Harvard Business Review (May 30, 2024) (hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams).

information like addresses, account numbers, and birthdates, bad actors can more realistically impersonate a victim's bank, a government office, or even a member of their own family in an attempt to gain control of accounts or induce fraudulent payments.⁵

Fraudsters can use AI models to craft and then send convincing email, text, and phone call scams at an industrial scale.⁶ As Deloitte's Center for Financial Services described, "fake content has never been easier to create – or harder to catch."⁷ In one growing type of scam, criminals have called victims using AI-generated voice clones of their loved ones and pretended to need money after an emergency.⁸ For instance, in June 2025, a New York man was sentenced to prison for his role in a high-tech, "elaborate grandparent scam" in which he stole around \$20,000 from three New Hampshire families after convincing them that their loved ones were in trouble.⁹ According to the *Union Leader*, "[v]ictims say the scam involved the use of artificial intelligence mimicking a loved one's voice to trick them into turning over money to bail that person out of jail."¹⁰ Research has shown that these calls can now be fully automated with the use of interactive AI models capable of carrying on real-time conversations, allowing scammers to potentially reach many people at little cost – but massive harm to victims.¹¹ Bad actors are

⁵ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁶ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/); U.S. PIRG Education Fund, *A Dark Side of AI: Text and Phone Scams Are Getting Worse* (Nov. 14, 2025) (pirg.org/edfund/articles/a-dark-side-of-ai-text-and-phone-scams-are-getting-worse/).

⁷ Deloitte's Center for Financial Services, *Generative AI is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking* (May 29, 2024) (www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html).

⁸ Federal Trade Commission, *Scammers Use AI to Enhance Their Family Emergency Schemes* (Mar. 20, 2023) (consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁹ *Bronx Man Gets 4 To 8 Years for Role in 'Grandparent Scam,'* New Hampshire Union Leader (Sept. 24, 2025) (www.unionleader.com/news/courts/bronx-man-gets-4-to-8-years-for-role-in-grandparent-scam/article_58a05fcd-5edd-4d76-8dae-8ccd3937dcac.html).

¹⁰ *Id.*

¹¹ *Voice-Enabled AI Agents Can Automate Everything, Even Your Phone Scams*, The Register (Oct. 24, 2024) (www.theregister.com/2024/10/24/openai_realtime_api_phone_scam/).

using this same voice-altering technology to imitate employees of government agencies, banks, utilities, and other offices.¹²

Impersonation scams are also executed over email, and AI has made fraudulent email messages “so good you simply cannot tell” that they are fake.¹³ All kinds of organizations have been impersonated in email phishing attacks, including federal agencies.¹⁴ The Department of Veterans Affairs (VA), for example, has warned veterans of emails that look like official communications and seek to extract benefits and personal information.¹⁵ The VA has also specifically spotlighted how scammers are using generative AI to target veterans in these fraudulent email campaigns and other scams.¹⁶ The Social Security Administration (SSA) Office of the Inspector General has also issued multiple alerts this year warning beneficiaries to look out for realistic emails that purport to be from SSA and try to lure users to a fraudulent website or trick them into downloading malicious software.¹⁷

Anthropic rightly prohibits the use of its products for fraudulent and unlawful activity, including specific actions commonly associated with scams, such as “[g]enerat[ing] content for fraudulent activities, schemes, scams, phishing, or malware.”¹⁸ AI companies, however, have

¹² *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert).

¹³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

¹⁴ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers); Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/).

¹⁵ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers).

¹⁶ Department of Veterans Affairs, *Protecting Yourself and Your Benefits from Cyber & Artificial Intelligence (AI) Threats* (Apr. 29, 2025) (department.va.gov/privacy/fact-sheet/protecting-yourself-and-your-benefits-from-cyber-artificial-intelligence-ai-threats/).

¹⁷ Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/); Office of the Inspector General, Social Security Administration, *Press Release: Beware of Scam Emails Asking to Download Statements* (Apr. 1, 2025) (oig.ssa.gov/scam-alerts/2025-04-01-beware-of-scam-emails-asking-to-download-statements/).

¹⁸ Anthropic, Usage Policy (www.anthropic.com/legal/aup) (accessed Dec. 5, 2025).

reportedly faced challenges in preventing the misuse of their technology.¹⁹ For example, users have tricked AI programs into performing prohibited tasks simply through the “clever phrasing” of a prompt, according to a Fox News report, and *Scientific American* has similarly reported that researchers have “repeatedly demonstrated ways to get around” measures designed to prevent users from using AI programs in harmful ways.²⁰ In addition, AI systems are often trained on sensitive data, with one recent study uncovering credit card and passport information in a widely used training set.²¹ Research has shown that “these systems can memorize the data and then expose it to other users as part of the outputs.”²² In the hands of a fraudster, this sensitive information could be used to develop a highly personalized scam or to steal someone’s identity outright.²³ These urgent and complex challenges call for collaboration between the private and public sectors, and we want to work together to find solutions to protect Americans. Accordingly, we seek responses to the requests below to better understand Anthropic’s operations and commitment to preventing AI-driven scams.

The following questions pertain to all iterations of all generative AI services and models that Anthropic has released to the public, including those available only in limited or beta release.

1. Describe Anthropic’s overall strategy to prevent the misuse of its technologies for scams and fraud – two activities that are in violation of its usage policies.
 - a. What alternative fraud-prevention strategies has your company considered or tested?

¹⁹ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²⁰ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²¹ Rachel Hong et al., *A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset*, arXiv:2506.17185 [cs.CR] (June 20, 2025) (arxiv.org/html/2506.17185v1#bib).

²² Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era*, Stanford University Human-Centered Artificial Intelligence (Feb. 2024) (hai.stanford.edu/assets/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf).

²³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, *Forbes* (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

- b. How much does the company invest to prevent scams and fraud?
 - c. How many people does Anthropic employ, either as employees of the company or outside consultants?
 - d. How many people does Anthropic employ, either as employees of the company or outside consultants, to work on preventing scams?
- 2. Does Anthropic monitor for user activity that violates the company's usage policies prohibiting the use of its models for scams and fraud?
 - a. Why or why not?
- 3. If Anthropic does track violations of its policies prohibiting the use of its services to scam or defraud, please provide annual data for each Anthropic model and version regarding:
 - a. The number of instances in which any proactive detection tools used by Anthropic identified content or inputs that violated policies prohibiting the use of its services to scam or defraud;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - b. The number of instances in which Anthropic was alerted to potential violations of its policies prohibiting the use of its services to scam or defraud by user reports or external notices;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - c. The number of individual accounts and individual users whose accounts Anthropic has deactivated and/or suspended for violating its policies against scams and fraud.
 - i. Please provide the number of these deactivations and suspensions that occurred following human review, if any. Please explain the level of review Anthropic requires prior to these actions, in general, and the rationale for these procedures.
 - d. The number of individual accounts and individual users against which Anthropic has taken action short of account suspensions or terminations, including through warnings or account restrictions, broken down by type of action;
 - i. Please provide the number of actions that were taken following human review, if any. Please explain the level of review Anthropic requires

prior to these actions, in general, and the rationale for these procedures.

- e. A breakdown of the severity and nature of the activities that Anthropic determined were in violation of its policies against scams and fraud (i.e., the specific usage policy or policies violated);
 - i. Of these violations, please state the number Anthropic initially identified via proactive detection, user reports or external notices, and human review, respectively.
4. Please detail the steps Anthropic takes to authenticate users, specifically:
- a. Please describe what, if any, know-your-customer information Anthropic collects, including but not limited to user names, emails, phone numbers, credit card information, IP addresses, and physical location. Please describe efforts by Anthropic, if any, to authenticate these user details.
 - b. Please detail the steps that the company takes to ensure that banned users are prohibited from creating new Anthropic accounts or operating additional existing accounts under a different name and/or with fraudulent credentials.
 - c. If Anthropic detects someone whom the platform banned for illegal activity attempting to sign up for a new account or otherwise operate another account, does Anthropic alert law enforcement? If so, under what circumstances does Anthropic make these alerts? If not, why not?
5. Does Anthropic attempt to determine if users and/or accounts that violate its policies regarding scams and fraud are performing these activities in coordination with other parties?
- a. For example, does Anthropic look for evidence or patterns that suggest violators may be working as part of a broader criminal network or on behalf of a foreign government?
 - b. Similarly, does Anthropic try to determine if these user accounts or activities belong to or are otherwise tied to individuals the United States has designated as threats, including Specially Designated Nationals, other sanctioned individuals, individuals designated as foreign or domestic terrorists, or wanted criminals?
 - i. If so, how many such user accounts has Anthropic identified?
 - ii. If so, does Anthropic alert law enforcement to such instances? If so, under what circumstances does Anthropic make these alerts? If not, why not?

6. Please describe Anthropic's policies and procedures regarding location-specific access restrictions, also known as "geofencing," including with respect to restrictions designed to prevent the use of its products in known hotspots for scam activity.
7. How many outputs have Anthropic's models generated for users that violated the company's own usage policies with respect to scams and/or fraud, including but not limited to providing written information or performing functions that were impermissible under the company's rules?
 - a. What share of these model outputs were in response to user inputs that violated the company's usage policies regarding scams and/or fraud?
 - b. What share of these model outputs were in response to user inputs considered acceptable under the company's usage policies?
8. Please describe any measures Anthropic has in place to monitor for and/or prevent the user activities listed below that can be used to facilitate scams and/or fraud.
 - a. Drafting content, including emails, text messages, or scripts for audio or video materials, that suggest that the user is a trusted entity like a bank or government agency;
 - b. Drafting or preparing recipient lists for emails, text messages, or other content – particularly to a high volume of recipients – that ask the recipient to make a payment or provide sensitive personal information, such as bank account numbers, identifying information, or login credentials, or direct the recipient to a website designed to collect such a payment or personal information;
 - c. Drafting or preparing recipient lists for emails or text messages that contain malware, viruses, or other malicious content designed to take over or otherwise corrupt the recipient's accounts, software, or hardware;
 - d. Inputs designed to facilitate the theft of passwords or login credentials; and
 - e. Inputs designed to bypass biometric account verification measures such as facial recognition.
 - f. If Anthropic monitors for the activities described in Requests 8 (a)–(f) above, how does it 1) do so while still respecting the privacy of its users and 2) differentiate between legitimate communications – such as a business sending a mass email promotion – and nefarious ones?
 - g. If Anthropic identifies the content described in Requests 8 (a)–(f), what actions, if any, does Anthropic take to alert law enforcement and/or any of the

individuals or entities that were impersonated or targeted? If Anthropic takes no action, please explain.

9. What steps does Anthropic take to prevent Personally Identifiable Information (PII) contained in illicit sources – such as from data breaches, hacks, or other leaks of sensitive and/or classified information from public or private sector sources – from being used as part of its training data?
 - a. Does Anthropic conduct regular training data audits to ensure that illicitly-sourced PII is not in its training data? If so, how often are such audits conducted?
 - b. Has Anthropic alerted government officials after discovering PII that was illegally sourced? If so, please describe the frequency of these alerts and the entities contacted.
10. Has Anthropic provided a way for an individual with information online to opt out of having their PII or other sensitive information included in the training content of its products?
11. For the users whose personal data is being used to train Anthropic's models, is the company detecting and/or removing any PII or other sensitive information in users' chat histories – including discussions of any sensitive financial information, medical history, etc., for any individual – from training material?
12. What steps does Anthropic take prior to a product release to ensure that its programs will not release PII? Have concerns over PII ever delayed Anthropic in releasing a product? If so, which products were delayed?
13. Can Anthropic detect if the programs it has released to the public share PII with other users, and if so, how? How does Anthropic monitor its programs in real time, if at all, to identify, prevent, and correct such incidents?
14. If an Anthropic product improperly shares PII, does the company inform the individuals whose data has been released?
15. Please describe any cooperation or coordination between Anthropic and law enforcement agencies at the international, federal, state, and/or local levels to prevent or investigate the use of its products for fraudulent purposes.
16. Please describe any cooperation or coordination between Anthropic and government agencies that are responsible for combating fraud, including but not limited to the Federal Trade Commission.

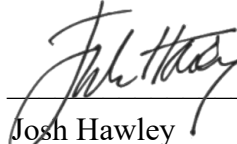
Mr. Dario Amodei
December 11, 2025
Page 9

Please provide your responses as soon as possible but in no event later than January 14, 2026. If you have any questions related to this request, please contact [REDACTED] of the [REDACTED] staff at [REDACTED]. Please send any official correspondence relating to this request to [REDACTED].

Sincerely,



Margaret Wood Hassan
Ranking Member
U.S. Congress Joint Economic
Committee



Josh Hawley
U.S. Senator

cc: David Schweikert
Chairman, Joint Economic Committee

Eric Schmitt
Vice Chairman, Joint Economic Committee

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mr. Sundar Pichai
Chief Executive Officer
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043

Dear Mr. Pichai:

We write today to request information on Google's efforts to implement sufficient protections to prevent scammers from misusing its technology. In recent years, global criminal networks have turned to AI to target more people with increasingly personalized and believable digital scams, contributing to a booming scam economy that is now a larger illicit industry than the global drug trade.¹ The FBI recorded \$16.6 billion in reported losses to suspected scams and other cybercrime in 2024 – up from \$3.5 billion in 2019, a more than 370 percent increase over the last five years.² The federal government has a responsibility to protect the American people from scams, but this effort requires an all-hands-on-deck approach across multiple industries. We hope we can work together productively on this important issue.

With advancements in AI, scams will continue to grow in sophistication, frequency, and impact.³ In the early phases of a scam, criminals can use generative AI to quickly identify and

¹ *Online Scams May Already Be as Big a Scourge as Illegal Drugs*, The Economist (Feb. 6, 2025) (www.economist.com/briefing/2025/02/06/online-scams-may-already-be-as-big-a-scourge-as-illegal-drugs); The Economist Group, *Press Release: Fraud Factories, Cyber Criminals and Corruption: The Economist's New Podcast, "Scam Inc", Uncovers a New, Global, Underground Economy Worth More Than Illicit Drug Trade* (Feb. 6, 2025) (www.economistgroup.com/press-centre/the-economist/fraud-factories-cyber-criminals-and-corruption-the-economists-new-podcast-scam-inc).

² FBI, *Internet Crime Report* (2024) (www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf); FBI, *Internet Crime Report* (2019) (www.ic3.gov/AnnualReport/Reports/2019_ic3Report.pdf).

³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

then collect details on their targets, enabling them to create tailor-made scams.⁴ Once armed with information like addresses, account numbers, and birthdates, bad actors can more realistically impersonate a victim's bank, a government office, or even a member of their own family in an attempt to gain control of accounts or induce fraudulent payments.⁵

Fraudsters can use AI models to craft and then send convincing email, text, and phone call scams at an industrial scale.⁶ As Deloitte's Center for Financial Services described, "fake content has never been easier to create – or harder to catch."⁷ In one growing type of scam, criminals have called victims using AI-generated voice clones of their loved ones and pretended to need money after an emergency.⁸ For instance, in June 2025, a New York man was sentenced to prison for his role in a high-tech, "elaborate grandparent scam" in which he stole around \$20,000 from three New Hampshire families after convincing them that their loved ones were in trouble.⁹ According to the *Union Leader*, "[v]ictims say the scam involved the use of artificial intelligence mimicking a loved one's voice to trick them into turning over money to bail that person out of jail."¹⁰ Research has shown that these calls can now be fully automated with the

⁴ Fredrik Heiding, Bruce Schneier, and Arun Vishwanath, *AI Will Increase the Quantity — and Quality — of Phishing Scams*, Harvard Business Review (May 30, 2024) (hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams).

⁵ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁶ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/); U.S. PIRG Education Fund, *A Dark Side of AI: Text and Phone Scams Are Getting Worse* (Nov. 14, 2025) (pirg.org/edfund/articles/a-dark-side-of-ai-text-and-phone-scams-are-getting-worse/).

⁷ Deloitte's Center for Financial Services, *Generative AI is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking* (May 29, 2024) (www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html).

⁸ Federal Trade Commission, *Scammers Use AI to Enhance Their Family Emergency Schemes* (Mar. 20, 2023) (consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁹ *Bronx Man Gets 4 To 8 Years for Role in 'Grandparent Scam,'* New Hampshire Union Leader (Sept. 24, 2025) (www.unionleader.com/news/courts/bronx-man-gets-4-to-8-years-for-role-in-grandparent-scam/article_58a05fcd-5edd-4d76-8dae-8ccd3937dcac.html).

¹⁰ *Id.*

use of interactive AI models capable of carrying on real-time conversations, allowing scammers to potentially reach many people at little cost – but massive harm to victims.¹¹ Bad actors are using this same voice-altering technology to imitate employees of government agencies, banks, utilities, and other offices.¹²

Impersonation scams are also executed over email, and AI has made fraudulent email messages “so good you simply cannot tell” that they are fake.¹³ All kinds of organizations have been impersonated in email phishing attacks, including federal agencies.¹⁴ The Department of Veterans Affairs (VA), for example, has warned veterans of emails that look like official communications and seek to extract benefits and personal information.¹⁵ The VA has also specifically spotlighted how scammers are using generative AI to target veterans in these fraudulent email campaigns and other scams.¹⁶ The Social Security Administration (SSA) Office of the Inspector General has also issued multiple alerts this year warning beneficiaries to look out for realistic emails that purport to be from SSA and try to lure users to a fraudulent website or trick them into downloading malicious software.¹⁷

¹¹ *Voice-Enabled AI Agents Can Automate Everything, Even Your Phone Scams*, The Register (Oct. 24, 2024) (www.theregister.com/2024/10/24/openai_realtime_api_phone_scam/).

¹² *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert).

¹³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

¹⁴ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers); Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/).

¹⁵ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers).

¹⁶ Department of Veterans Affairs, *Protecting Yourself and Your Benefits from Cyber & Artificial Intelligence (AI) Threats* (Apr. 29, 2025) (department.va.gov/privacy/fact-sheet/protecting-yourself-and-your-benefits-from-cyber-artificial-intelligence-ai-threats/).

¹⁷ Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/); Office of the Inspector General, Social Security Administration, *Press Release: Beware of Scam Emails Asking to Download Statements* (Apr. 1, 2025) (oig.ssa.gov/scam-alerts/2025-04-01-beware-of-scam-emails-asking-to-download-statements/).

Google rightly prohibits the use of its products for fraud and scams.¹⁸ AI companies, however, have reportedly faced challenges in preventing the misuse of their technology.¹⁹ For example, users have tricked AI programs into performing prohibited tasks simply through the “clever phrasing” of a prompt, according to a Fox News report, and *Scientific American* has similarly reported that researchers have “repeatedly demonstrated ways to get around” measures designed to prevent users from using AI programs in harmful ways.²⁰ In addition, AI systems are often trained on sensitive data, with one recent study uncovering credit card and passport information in a widely used training set.²¹ Research has shown that “these systems can memorize the data and then expose it to other users as part of the outputs.”²² In the hands of a fraudster, this sensitive information could be used to develop a highly personalized scam or to steal someone’s identity outright.²³ These urgent and complex challenges call for collaboration between the private and public sectors, and we want to work together to find solutions to protect Americans. Accordingly, we seek responses to the requests below to better understand Google’s operations and commitment to preventing AI-driven scams.

The following questions pertain to all iterations of all generative AI services and models that Google has released to the public, including those available only in limited or beta release.

1. Describe Google’s overall strategy to prevent the misuse of its technologies for scams and fraud – two activities that are in violation of its usage policies.

¹⁸ Google, Generative AI Prohibited Use Policy (policies.google.com/terms/generative-ai/use-policy) (accessed Dec. 4, 2025).

¹⁹ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²⁰ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²¹ Rachel Hong et al., *A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset*, arXiv:2506.17185 [cs.CR] (June 20, 2025) (arxiv.org/html/2506.17185v1#bib).

²² Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era*, Stanford University Human-Centered Artificial Intelligence (Feb. 2024) (hai.stanford.edu/assets/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf).

²³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, *Forbes* (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

- a. What alternative fraud-prevention strategies has your company considered or tested?
 - b. How much does the company invest to prevent scams and fraud?
 - c. How many people does Google employ, either as employees of the company or outside consultants?
 - d. How many people does Google employ, either as employees of the company or outside consultants, to work on preventing scams?
2. Does Google monitor for user activity that violates the company's usage policies prohibiting the use of its models for scams and fraud?
 - a. Why or why not?
3. If Google does track violations of its policies prohibiting the use of its services to scam or defraud, please provide annual data for each Google model and version regarding:
 - a. The number of instances in which any proactive detection tools used by Google identified content or inputs that violated policies prohibiting the use of its services to scam or defraud;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - b. The number of instances in which Google was alerted to potential violations of its policies prohibiting the use of its services to scam or defraud by user reports or external notices;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - c. The number of individual accounts and individual users whose accounts Google has deactivated and/or suspended for violating its policies against scams and fraud.
 - i. Please provide the number of these deactivations and suspensions that occurred following human review, if any. Please explain the level of review Google requires prior to these actions, in general, and the rationale for these procedures.
 - d. The number of individual accounts and individual users against which Google has taken action short of account suspensions or terminations, including through warnings or account restrictions, broken down by type of action;

- i. Please provide the number of actions that were taken following human review, if any. Please explain the level of review Google requires prior to these actions, in general, and the rationale for these procedures.
 - e. A breakdown of the severity and nature of the activities that Google determined were in violation of its policies against scams and fraud (i.e., the specific usage policy or policies violated);
 - i. Of these violations, please state the number Google initially identified via proactive detection, user reports or external notices, and human review, respectively.
- 4. Please detail the steps Google takes to authenticate users, specifically:
 - a. Please describe what, if any, know-your-customer information Google collects, including but not limited to user names, emails, phone numbers, credit card information, IP addresses, and physical location. Please describe efforts by Google, if any, to authenticate these user details.
 - b. Please detail the steps that the company takes to ensure that banned users are prohibited from creating new Google accounts or operating additional existing accounts under a different name and/or with fraudulent credentials.
 - c. If Google detects someone whom the platform banned for illegal activity attempting to sign up for a new account or otherwise operate another account, does Google alert law enforcement? If so, under what circumstances does Google make these alerts? If not, why not?
- 5. Does Google attempt to determine if users and/or accounts that violate its policies regarding scams and fraud are performing these activities in coordination with other parties?
 - a. For example, does Google look for evidence or patterns that suggest violators may be working as part of a broader criminal network or on behalf of a foreign government?
 - b. Similarly, does Google try to determine if these user accounts or activities belong to or are otherwise tied to individuals the United States has designated as threats, including Specially Designated Nationals, other sanctioned individuals, individuals designated as foreign or domestic terrorists, or wanted criminals?
 - i. If so, how many such user accounts has Google identified?
 - ii. If so, does Google alert law enforcement to such instances? If so, under what circumstances does Google make these alerts? If not, why not?

6. Please describe Google's policies and procedures regarding location-specific access restrictions, also known as "geofencing," including with respect to restrictions designed to prevent the use of its products in known hotspots for scam activity.
7. How many outputs have Google's models generated for users that violated the company's own usage policies with respect to scams and/or fraud, including but not limited to providing written information, producing visual content, or performing functions that were impermissible under the company's rules?
 - a. What share of these model outputs were in response to user inputs that violated the company's usage policies regarding scams and/or fraud?
 - b. What share of these model outputs were in response to user inputs considered acceptable under the company's usage policies?
8. Please describe any measures Google has in place to monitor for and/or prevent the user activities listed below that can be used to facilitate scams and/or fraud.
 - a. Recreating logos of an entity, including banks and government offices, without permission;
 - b. Drafting emails, text messages, videos, or voice messages, including ones that utilize AI-generated voice technology, that suggest that the user is a trusted entity like a bank or government agency;
 - c. Drafting or preparing recipient lists for emails, text messages, videos, or voice messages – particularly to a high volume of recipients – that ask the recipient to make a payment or provide sensitive personal information, such as bank account numbers, identifying information, or login credentials, or direct the recipient to a website designed to collect such a payment or personal information;
 - d. Drafting or preparing recipient lists for emails or text messages that contain malware, viruses, or other malicious content designed to take over or otherwise corrupt the recipient's accounts, software, or hardware;
 - e. Inputs designed to facilitate the theft of passwords or login credentials; and
 - f. Inputs designed to bypass biometric account verification measures such as facial recognition.
 - g. If Google monitors for the activities described in Requests 8 (a)–(f) above, how does it 1) do so while still respecting the privacy of its users and 2) differentiate between legitimate communications – such as a business sending a mass email promotion – and nefarious ones?

- h. If Google identifies the content described in Requests 8 (a)–(f), what actions, if any, does Google take to alert law enforcement and/or any of the individuals or entities that were impersonated or targeted? If Google takes no action, please explain.
- 9. What steps does Google take to prevent Personally Identifiable Information (PII) contained in illicit sources – such as from data breaches, hacks, or other leaks of sensitive and/or classified information from public or private sector sources – from being used as part of its training data?
 - a. Does Google conduct regular training data audits to ensure that illicitly-sourced PII is not in its training data? If so, how often are such audits conducted?
 - b. Has Google alerted government officials after discovering PII that was illegally sourced? If so, please describe the frequency of these alerts and the entities contacted.
- 10. Has Google provided a way for an individual with information online to opt out of having their PII or other sensitive information included in the training content of its products?
- 11. For the users whose personal data is being used to train Google’s models, is the company detecting and/or removing any PII or other sensitive information in users’ chat histories – including discussions of any sensitive financial information, medical history, etc., for any individual – from training material?
- 12. What steps does Google take prior to a product release to ensure that its programs will not release PII? Have concerns over PII ever delayed Google in releasing a product? If so, which products were delayed?
- 13. Can Google detect if the programs it has released to the public share PII with other users, and if so, how? How does Google monitor its programs in real time, if at all, to identify, prevent, and correct such incidents?
- 14. If a Google product improperly shares PII, does the company inform the individuals whose data has been released?
- 15. Please describe any cooperation or coordination between Google and law enforcement agencies at the international, federal, state, and/or local levels to prevent or investigate the use of its products for fraudulent purposes.

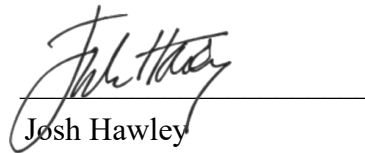
16. Please describe any cooperation or coordination between Google and government agencies that are responsible for combating fraud, including but not limited to the Federal Trade Commission.

Please provide your responses as soon as possible but in no event later than January 14, 2026. If you have any questions related to this request, please contact [REDACTED] of the Joint Economic Committee staff at [REDACTED]. Please send any official correspondence relating to this request to [REDACTED].

Sincerely,



Margaret Wood Hassan
Ranking Member
U.S. Congress Joint Economic
Committee



Josh Hawley
U.S. Senator

cc: David Schweikert
Chairman, Joint Economic Committee

Eric Schmitt
Vice Chairman, Joint Economic Committee

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mr. Mark Zuckerberg
Chairman and Chief Executive Officer
Meta
1 Meta Way
Menlo Park, CA 94025

Dear Mr. Zuckerberg:

We write today to request information on Meta's efforts to implement sufficient protections to prevent scammers from misusing its technology. In recent years, global criminal networks have turned to AI to target more people with increasingly personalized and believable digital scams, contributing to a booming scam economy that is now a larger illicit industry than the global drug trade.¹ The FBI recorded \$16.6 billion in reported losses to suspected scams and other cybercrime in 2024 – up from \$3.5 billion in 2019, a more than 370 percent increase over the last five years.² The federal government has a responsibility to protect the American people from scams, but this effort requires an all-hands-on-deck approach across multiple industries. We hope we can work together productively on this important issue.

With advancements in AI, scams will continue to grow in sophistication, frequency, and impact.³ In the early phases of a scam, criminals can use generative AI to quickly identify and

¹ *Online Scams May Already Be as Big a Scourge as Illegal Drugs*, The Economist (Feb. 6, 2025) (www.economist.com/briefing/2025/02/06/online-scams-may-already-be-as-big-a-scourge-as-illegal-drugs); The Economist Group, *Press Release: Fraud Factories, Cyber Criminals and Corruption: The Economist's New Podcast, "Scam Inc", Uncovers a New, Global, Underground Economy Worth More Than Illicit Drug Trade* (Feb. 6, 2025) (www.economistgroup.com/press-centre/the-economist/fraud-factories-cyber-criminals-and-corruption-the-economists-new-podcast-scam-inc).

² FBI, *Internet Crime Report* (2024) (www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf); FBI, *Internet Crime Report* (2019) (www.ic3.gov/AnnualReport/Reports/2019_ic3Report.pdf).

³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

then collect details on their targets, enabling them to create tailor-made scams.⁴ Once armed with information like addresses, account numbers, and birthdates, bad actors can more realistically impersonate a victim's bank, a government office, or even a member of their own family in an attempt to gain control of accounts or induce fraudulent payments.⁵

Fraudsters can use AI models to craft and then send convincing email, text, and phone call scams at an industrial scale.⁶ As Deloitte's Center for Financial Services described, "fake content has never been easier to create – or harder to catch."⁷ In one growing type of scam, criminals have called victims using AI-generated voice clones of their loved ones and pretended to need money after an emergency.⁸ For instance, in June 2025, a New York man was sentenced to prison for his role in a high-tech, "elaborate grandparent scam" in which he stole around \$20,000 from three New Hampshire families after convincing them that their loved ones were in trouble.⁹ According to the *Union Leader*, "[v]ictims say the scam involved the use of artificial intelligence mimicking a loved one's voice to trick them into turning over money to bail that person out of jail."¹⁰ Research has shown that these calls can now be fully automated with the

⁴ Fredrik Heiding, Bruce Schneier, and Arun Vishwanath, *AI Will Increase the Quantity — and Quality — of Phishing Scams*, Harvard Business Review (May 30, 2024) (hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams).

⁵ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁶ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/); U.S. PIRG Education Fund, *A Dark Side of AI: Text and Phone Scams Are Getting Worse* (Nov. 14, 2025) (pirg.org/edfund/articles/a-dark-side-of-ai-text-and-phone-scams-are-getting-worse/).

⁷ Deloitte's Center for Financial Services, *Generative AI is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking* (May 29, 2024) (www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html).

⁸ Federal Trade Commission, *Scammers Use AI to Enhance Their Family Emergency Schemes* (Mar. 20, 2023) (consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁹ *Bronx Man Gets 4 To 8 Years for Role in 'Grandparent Scam,'* New Hampshire Union Leader (Sept. 24, 2025) (www.unionleader.com/news/courts/bronx-man-gets-4-to-8-years-for-role-in-grandparent-scam/article_58a05fcd-5edd-4d76-8dae-8ccd3937dcac.html).

¹⁰ *Id.*

use of interactive AI models capable of carrying on real-time conversations, allowing scammers to potentially reach many people at little cost – but massive harm to victims.¹¹ Bad actors are using this same voice-altering technology to imitate employees of government agencies, banks, utilities, and other offices.¹²

Impersonation scams are also executed over email, and AI has made fraudulent email messages “so good you simply cannot tell” that they are fake.¹³ All kinds of organizations have been impersonated in email phishing attacks, including federal agencies.¹⁴ The Department of Veterans Affairs (VA), for example, has warned veterans of emails that look like official communications and seek to extract benefits and personal information.¹⁵ The VA has also specifically spotlighted how scammers are using generative AI to target veterans in these fraudulent email campaigns and other scams.¹⁶ The Social Security Administration (SSA) Office of the Inspector General has also issued multiple alerts this year warning beneficiaries to look out for realistic emails that purport to be from SSA and try to lure users to a fraudulent website or trick them into downloading malicious software.¹⁷

¹¹ *Voice-Enabled AI Agents Can Automate Everything, Even Your Phone Scams*, The Register (Oct. 24, 2024) (www.theregister.com/2024/10/24/openai_realtime_api_phone_scam/).

¹² *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert).

¹³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

¹⁴ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers); Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/).

¹⁵ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers).

¹⁶ Department of Veterans Affairs, *Protecting Yourself and Your Benefits from Cyber & Artificial Intelligence (AI) Threats* (Apr. 29, 2025) (department.va.gov/privacy/fact-sheet/protecting-yourself-and-your-benefits-from-cyber-artificial-intelligence-ai-threats/).

¹⁷ Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/); Office of the Inspector General, Social Security Administration, *Press Release: Beware of Scam Emails Asking to Download Statements* (Apr. 1, 2025) (oig.ssa.gov/scam-alerts/2025-04-01-beware-of-scam-emails-asking-to-download-statements/).

Meta rightly prohibits the use of its products for fraudulent and unlawful activity, including specifically “activity related to ... scams, phishing or malware.”¹⁸ AI companies, however, have reportedly faced challenges in preventing the misuse of their technology.¹⁹ For example, users have tricked AI programs into performing prohibited tasks simply through the “clever phrasing” of a prompt, according to a Fox News report, and *Scientific American* has similarly reported that researchers have “repeatedly demonstrated ways to get around” measures designed to prevent users from using AI programs in harmful ways.²⁰ In addition, AI systems are often trained on sensitive data, with one recent study uncovering credit card and passport information in a widely used training set.²¹ Research has shown that “these systems can memorize the data and then expose it to other users as part of the outputs.”²² In the hands of a fraudster, this sensitive information could be used to develop a highly personalized scam or to steal someone’s identity outright.²³ These urgent and complex challenges call for collaboration between the private and public sectors, and we want to work together to find solutions to protect Americans. Accordingly, we seek responses to the requests below to better understand Meta’s operations and commitment to preventing AI-driven scams.

The following questions pertain to all iterations of all generative AI services and models that Meta has released to the public, including those available only in limited or beta release.

¹⁸ Facebook, Meta AIs Terms of Service (www.facebook.com/legal/ai-terms/) (accessed Dec. 4, 2025).

¹⁹ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²⁰ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²¹ Rachel Hong et al., *A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset*, arXiv:2506.17185 [cs.CR] (June 20, 2025) (arxiv.org/html/2506.17185v1#bib).

²² Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era*, Stanford University Human-Centered Artificial Intelligence (Feb. 2024) (hai.stanford.edu/assets/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf).

²³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, *Forbes* (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

1. Describe Meta's overall strategy to prevent the misuse of its technologies for scams and fraud – two activities that are in violation of its usage policies.
 - a. What alternative fraud-prevention strategies has your company considered or tested?
 - b. How much does the company invest to prevent scams and fraud?
 - c. How many people does Meta employ, either as employees of the company or outside consultants?
 - d. How many people does Meta employ, either as employees of the company or outside consultants, to work on preventing scams?
2. Does Meta monitor for user activity that violates the company's usage policies prohibiting the use of its models for scams and fraud?
 - a. Why or why not?
3. If Meta does track violations of its policies prohibiting the use of its services to scam or defraud, please provide annual data for each Meta model and version regarding:
 - a. The number of instances in which any proactive detection tools used by Meta identified content or inputs that violated policies prohibiting the use of its services to scam or defraud;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - b. The number of instances in which Meta was alerted to potential violations of its policies prohibiting the use of its services to scam or defraud by user reports or external notices;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - c. The number of individual accounts and individual users whose accounts Meta has deactivated and/or suspended for violating its policies against scams and fraud.
 - i. Please provide the number of these deactivations and suspensions that occurred following human review, if any. Please explain the level of review Meta requires prior to these actions, in general, and the rationale for these procedures.
 - d. The number of individual accounts and individual users against which Meta has taken action short of account suspensions or terminations, including through warnings or account restrictions, broken down by type of action;

- i. Please provide the number of actions that were taken following human review, if any. Please explain the level of review Meta requires prior to these actions, in general, and the rationale for these procedures.
 - e. A breakdown of the severity and nature of the activities that Meta determined were in violation of its policies against scams and fraud (i.e., the specific usage policy or policies violated);
 - i. Of these violations, please state the number Meta initially identified via proactive detection, user reports or external notices, and human review, respectively.
- 4. Please detail the steps Meta takes to authenticate users, specifically:
 - a. Please describe what, if any, know-your-customer information Meta collects, including but not limited to user names, emails, phone numbers, credit card information, IP addresses, and physical location. Please describe efforts by Meta, if any, to authenticate these user details.
 - b. Please detail the steps that the company takes to ensure that banned users are prohibited from creating new Meta accounts or operating additional existing accounts under a different name and/or with fraudulent credentials.
 - c. If Meta detects someone whom the platform banned for illegal activity attempting to sign up for a new account or otherwise operate another account, does Meta alert law enforcement? If so, under what circumstances does Meta make these alerts? If not, why not?
- 5. Does Meta attempt to determine if users and/or accounts that violate its policies regarding scams and fraud are performing these activities in coordination with other parties?
 - a. For example, does Meta look for evidence or patterns that suggest violators may be working as part of a broader criminal network or on behalf of a foreign government?
 - b. Similarly, does Meta try to determine if these user accounts or activities belong to or are otherwise tied to individuals the United States has designated as threats, including Specially Designated Nationals, other sanctioned individuals, individuals designated as foreign or domestic terrorists, or wanted criminals?
 - i. If so, how many such user accounts has Meta identified?
 - ii. If so, does Meta alert law enforcement to such instances? If so, under what circumstances does Meta make these alerts? If not, why not?

6. Please describe Meta's policies and procedures regarding location-specific access restrictions, also known as "geofencing," including with respect to restrictions designed to prevent the use of its products in known hotspots for scam activity.
7. How many outputs have Meta's models generated for users that violated the company's own usage policies with respect to scams and/or fraud, including but not limited to providing written information, producing visual content, or performing functions that were impermissible under the company's rules?
 - a. What share of these model outputs were in response to user inputs that violated the company's usage policies regarding scams and/or fraud?
 - b. What share of these model outputs were in response to user inputs considered acceptable under the company's usage policies?
8. Please describe any measures Meta has in place to monitor for and/or prevent the user activities listed below that can be used to facilitate scams and/or fraud.
 - a. Recreating logos of an entity, including banks and government offices, without permission;
 - b. Drafting emails, text messages, videos, or voice messages, including ones that utilize AI-generated voice technology, that suggest that the user is a trusted entity like a bank or government agency;
 - c. Drafting or preparing recipient lists for emails, text messages, videos, or voice messages – particularly to a high volume of recipients – that ask the recipient to make a payment or provide sensitive personal information, such as bank account numbers, identifying information, or login credentials, or direct the recipient to a website designed to collect such a payment or personal information;
 - d. Drafting or preparing recipient lists for emails or text messages that contain malware, viruses, or other malicious content designed to take over or otherwise corrupt the recipient's accounts, software, or hardware;
 - e. Inputs designed to facilitate the theft of passwords or login credentials; and
 - f. Inputs designed to bypass biometric account verification measures such as facial recognition.
 - g. If Meta monitors for the activities described in Requests 8 (a)–(f) above, how does it 1) do so while still respecting the privacy of its users and 2) differentiate between legitimate communications – such as a business sending a mass email promotion – and nefarious ones?

- h. If Meta identifies the content described in Requests 8 (a)–(f), what actions, if any, does Meta take to alert law enforcement and/or any of the individuals or entities that were impersonated or targeted? If Meta takes no action, please explain.
9. What steps does Meta take to prevent Personally Identifiable Information (PII) contained in illicit sources – such as from data breaches, hacks, or other leaks of sensitive and/or classified information from public or private sector sources – from being used as part of its training data?
 - a. Does Meta conduct regular training data audits to ensure that illicitly-sourced PII is not in its training data? If so, how often are such audits conducted?
 - b. Has Meta alerted government officials after discovering PII that was illegally sourced? If so, please describe the frequency of these alerts and the entities contacted.
10. Has Meta provided a way for an individual with information online to opt out of having their PII or other sensitive information included in the training content of its products?
11. For the users whose personal data is being used to train Meta’s models, is the company detecting and/or removing any PII or other sensitive information – including discussions of any sensitive financial information, medical history, etc., for any individual – from training material?
12. What steps does Meta take prior to a product release to ensure that its programs will not release PII? Have concerns over PII ever delayed Meta in releasing a product? If so, which products were delayed?
13. Can Meta detect if the programs it has released to the public share PII with other users, and if so, how? How does Meta monitor its programs in real time, if at all, to identify, prevent, and correct such incidents?
14. If a Meta product improperly shares PII, does the company inform the individuals whose data has been released?
15. Please describe any cooperation or coordination between Meta and law enforcement agencies at the international, federal, state, and/or local levels to prevent or investigate the use of its products for fraudulent purposes.
16. Please describe any cooperation or coordination between Meta and government agencies that are responsible for combating fraud, including but not limited to the Federal Trade Commission.

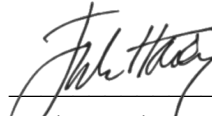
Mr. Mark Zuckerberg
December 11, 2025
Page 9

Please provide your responses as soon as possible but in no event later than January 14, 2026. If you have any questions related to this request, please contact [REDACTED] of the Joint Economic Committee staff at [REDACTED]. Please send any official correspondence relating to this request to [REDACTED] and [REDACTED].

Sincerely,



Margaret Wood Hassan
Ranking Member
U.S. Congress Joint Economic
Committee



Josh Hawley
U.S. Senator

cc: David Schweikert
Chairman, Joint Economic Committee

Eric Schmitt
Vice Chairman, Joint Economic Committee

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mr. Satya Nadella
Chief Executive Officer
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Dear Mr. Nadella:

We write today to request information on Microsoft's efforts to implement sufficient protections to prevent scammers from misusing its technology. In recent years, global criminal networks have turned to AI to target more people with increasingly personalized and believable digital scams, contributing to a booming scam economy that is now a larger illicit industry than the global drug trade.¹ The FBI recorded \$16.6 billion in reported losses to suspected scams and other cybercrime in 2024 – up from \$3.5 billion in 2019, a more than 370 percent increase over the last five years.² The federal government has a responsibility to protect the American people from scams, but this effort requires an all-hands-on-deck approach across multiple industries. We hope we can work together productively on this important issue.

With advancements in AI, scams will continue to grow in sophistication, frequency, and impact.³ In the early phases of a scam, criminals can use generative AI to quickly identify and

¹ *Online Scams May Already Be as Big a Scourge as Illegal Drugs*, The Economist (Feb. 6, 2025) (www.economist.com/briefing/2025/02/06/online-scams-may-already-be-as-big-a-scourge-as-illegal-drugs); The Economist Group, *Press Release: Fraud Factories, Cyber Criminals and Corruption: The Economist's New Podcast, "Scam Inc", Uncovers a New, Global, Underground Economy Worth More Than Illicit Drug Trade* (Feb. 6, 2025) (www.economistgroup.com/press-centre/the-economist/fraud-factories-cyber-criminals-and-corruption-the-economists-new-podcast-scam-inc).

² FBI, *Internet Crime Report* (2024) (www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf); FBI, *Internet Crime Report* (2019) (www.ic3.gov/AnnualReport/Reports/2019_ic3Report.pdf).

³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

then collect details on their targets, enabling them to create tailor-made scams.⁴ Once armed with information like addresses, account numbers, and birthdates, bad actors can more realistically impersonate a victim's bank, a government office, or even a member of their own family in an attempt to gain control of accounts or induce fraudulent payments.⁵

Fraudsters can use AI models to craft and then send convincing email, text, and phone call scams at an industrial scale.⁶ As Deloitte's Center for Financial Services described, "fake content has never been easier to create – or harder to catch."⁷ In one growing type of scam, criminals have called victims using AI-generated voice clones of their loved ones and pretended to need money after an emergency.⁸ For instance, in June 2025, a New York man was sentenced to prison for his role in a high-tech, "elaborate grandparent scam" in which he stole around \$20,000 from three New Hampshire families after convincing them that their loved ones were in trouble.⁹ According to the *Union Leader*, "[v]ictims say the scam involved the use of artificial intelligence mimicking a loved one's voice to trick them into turning over money to bail that person out of jail."¹⁰ Research has shown that these calls can now be fully automated with the

⁴ Fredrik Heiding, Bruce Schneier, and Arun Vishwanath, *AI Will Increase the Quantity — and Quality — of Phishing Scams*, Harvard Business Review (May 30, 2024) (hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams).

⁵ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁶ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/); U.S. PIRG Education Fund, *A Dark Side of AI: Text and Phone Scams Are Getting Worse* (Nov. 14, 2025) (pirg.org/edfund/articles/a-dark-side-of-ai-text-and-phone-scams-are-getting-worse/).

⁷ Deloitte's Center for Financial Services, *Generative AI is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking* (May 29, 2024) (www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html).

⁸ Federal Trade Commission, *Scammers Use AI to Enhance Their Family Emergency Schemes* (Mar. 20, 2023) (consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁹ *Bronx Man Gets 4 To 8 Years for Role in 'Grandparent Scam,'* New Hampshire Union Leader (Sept. 24, 2025) (www.unionleader.com/news/courts/bronx-man-gets-4-to-8-years-for-role-in-grandparent-scam/article_58a05fcd-5edd-4d76-8dae-8ccd3937dcac.html).

¹⁰ *Id.*

use of interactive AI models capable of carrying on real-time conversations, allowing scammers to potentially reach many people at little cost – but massive harm to victims.¹¹ Bad actors are using this same voice-altering technology to imitate employees of government agencies, banks, utilities, and other offices.¹²

Impersonation scams are also executed over email, and AI has made fraudulent email messages “so good you simply cannot tell” that they are fake.¹³ All kinds of organizations have been impersonated in email phishing attacks, including federal agencies.¹⁴ The Department of Veterans Affairs (VA), for example, has warned veterans of emails that look like official communications and seek to extract benefits and personal information.¹⁵ The VA has also specifically spotlighted how scammers are using generative AI to target veterans in these fraudulent email campaigns and other scams.¹⁶ The Social Security Administration (SSA) Office of the Inspector General has also issued multiple alerts this year warning beneficiaries to look out for realistic emails that purport to be from SSA and try to lure users to a fraudulent website or trick them into downloading malicious software.¹⁷

¹¹ *Voice-Enabled AI Agents Can Automate Everything, Even Your Phone Scams*, The Register (Oct. 24, 2024) (www.theregister.com/2024/10/24/openai_realtime_api_phone_scam/).

¹² *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert).

¹³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

¹⁴ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers); Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/).

¹⁵ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers).

¹⁶ Department of Veterans Affairs, *Protecting Yourself and Your Benefits from Cyber & Artificial Intelligence (AI) Threats* (Apr. 29, 2025) (department.va.gov/privacy/fact-sheet/protecting-yourself-and-your-benefits-from-cyber-artificial-intelligence-ai-threats/).

¹⁷ Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/); Office of the Inspector General, Social Security Administration, *Press Release: Beware of Scam Emails Asking to Download Statements* (Apr. 1, 2025) (oig.ssa.gov/scam-alerts/2025-04-01-beware-of-scam-emails-asking-to-download-statements/).

Microsoft rightly prohibits the use of its products for fraudulent and unlawful activity, including specific actions commonly associated with scams, such as sending spam, engaging in phishing, and generating or distributing malware.¹⁸ AI companies, however, have reportedly faced challenges in preventing the misuse of their technology.¹⁹ For example, users have tricked AI programs into performing prohibited tasks simply through the “clever phrasing” of a prompt, according to a Fox News report, and *Scientific American* has similarly reported that researchers have “repeatedly demonstrated ways to get around” measures designed to prevent users from using AI programs in harmful ways.²⁰ In addition, AI systems are often trained on sensitive data, with one recent study uncovering credit card and passport information in a widely used training set.²¹ Research has shown that “these systems can memorize the data and then expose it to other users as part of the outputs.”²² In the hands of a fraudster, this sensitive information could be used to develop a highly personalized scam or to steal someone’s identity outright.²³ These urgent and complex challenges call for collaboration between the private and public sectors, and we want to work together to find solutions to protect Americans. Accordingly, we seek responses to the requests below to better understand Microsoft’s operations and commitment to preventing AI-driven scams.

The following questions pertain to all iterations of all generative AI services and models that Microsoft has released to the public, including those available only in limited or beta release.

¹⁸ Microsoft, Microsoft Services Agreement (www.microsoft.com/en-us/servicesagreement#3_codeOfConduct) (accessed Dec. 5, 2025).

¹⁹ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²⁰ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²¹ Rachel Hong et al., *A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset*, arXiv:2506.17185 [cs.CR] (June 20, 2025) (arxiv.org/html/2506.17185v1#bib).

²² Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era*, Stanford University Human-Centered Artificial Intelligence (Feb. 2024) (hai.stanford.edu/assets/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf).

²³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, *Forbes* (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

1. Describe Microsoft's overall strategy to prevent the misuse of its technologies for scams and fraud – two activities that are in violation of its usage policies.
 - a. What alternative fraud-prevention strategies has your company considered or tested?
 - b. How much does the company invest to prevent scams and fraud?
 - c. How many people does Microsoft employ, either as employees of the company or outside consultants?
 - d. How many people does Microsoft employ, either as employees of the company or outside consultants, to work on preventing scams?
2. Does Microsoft monitor for user activity that violates the company's usage policies prohibiting the use of its models for scams and fraud?
 - a. Why or why not?
3. If Microsoft does track violations of its policies prohibiting the use of its services to scam or defraud, please provide annual data for each Microsoft model and version regarding:
 - a. The number of instances in which any proactive detection tools used by Microsoft identified content or inputs that violated policies prohibiting the use of its services to scam or defraud;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - b. The number of instances in which Microsoft was alerted to potential violations of its policies prohibiting the use of its services to scam or defraud by user reports or external notices;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - c. The number of individual accounts and individual users whose accounts Microsoft has deactivated and/or suspended for violating its policies against scams and fraud.
 - i. Please provide the number of these deactivations and suspensions that occurred following human review, if any. Please explain the level of review Microsoft requires prior to these actions, in general, and the rationale for these procedures.
 - d. The number of individual accounts and individual users against which Microsoft has taken action short of account suspensions or terminations,

including through warnings or account restrictions, broken down by type of action;

- i. Please provide the number of actions that were taken following human review, if any. Please explain the level of review Microsoft requires prior to these actions, in general, and the rationale for these procedures.
 - e. A breakdown of the severity and nature of the activities that Microsoft determined were in violation of its policies against scams and fraud (i.e., the specific usage policy or policies violated);
 - i. Of these violations, please state the number Microsoft initially identified via proactive detection, user reports or external notices, and human review, respectively.
4. Please detail the steps Microsoft takes to authenticate users, specifically:
 - a. Please describe what, if any, know-your-customer information Microsoft collects, including but not limited to user names, emails, phone numbers, credit card information, IP addresses, and physical location. Please describe efforts by Microsoft, if any, to authenticate these user details.
 - b. Please detail the steps that the company takes to ensure that banned users are prohibited from creating new Microsoft accounts or operating additional existing accounts under a different name and/or with fraudulent credentials.
 - c. If Microsoft detects someone whom the platform banned for illegal activity attempting to sign up for a new account or otherwise operate another account, does Microsoft alert law enforcement? If so, under what circumstances does Microsoft make these alerts? If not, why not?
5. Does Microsoft attempt to determine if users and/or accounts that violate its policies regarding scams and fraud are performing these activities in coordination with other parties?
 - a. For example, does Microsoft look for evidence or patterns that suggest violators may be working as part of a broader criminal network or on behalf of a foreign government?
 - b. Similarly, does Microsoft try to determine if these user accounts or activities belong to or are otherwise tied to individuals the United States has designated as threats, including Specially Designated Nationals, other sanctioned individuals, individuals designated as foreign or domestic terrorists, or wanted criminals?
 - i. If so, how many such user accounts has Microsoft identified?

- ii. If so, does Microsoft alert law enforcement to such instances? If so, under what circumstances does Microsoft make these alerts? If not, why not?
6. Please describe Microsoft's policies and procedures regarding location-specific access restrictions, also known as "geofencing," including with respect to restrictions designed to prevent the use of its products in known hotspots for scam activity.
7. How many outputs have Microsoft's models generated for users that violated the company's own usage policies with respect to scams and/or fraud, including but not limited to providing written information, producing visual content, or performing functions that were impermissible under the company's rules?
- a. What share of these model outputs were in response to user inputs that violated the company's usage policies regarding scams and/or fraud?
 - b. What share of these model outputs were in response to user inputs considered acceptable under the company's usage policies?
8. Please describe any measures Microsoft has in place to monitor for and/or prevent the user activities listed below that can be used to facilitate scams and/or fraud.
- a. Recreating logos of an entity, including banks and government offices, without permission;
 - b. Drafting emails, text messages, videos, or voice messages, including ones that utilize AI-generated voice technology, that suggest that the user is a trusted entity like a bank or government agency;
 - c. Drafting or preparing recipient lists for emails, text messages, videos, or voice messages – particularly to a high volume of recipients – that ask the recipient to make a payment or provide sensitive personal information, such as bank account numbers, identifying information, or login credentials, or direct the recipient to a website designed to collect such a payment or personal information;
 - d. Drafting or preparing recipient lists for emails or text messages that contain malware, viruses, or other malicious content designed to take over or otherwise corrupt the recipient's accounts, software, or hardware;
 - e. Inputs designed to facilitate the theft of passwords or login credentials; and
 - f. Inputs designed to bypass biometric account verification measures such as facial recognition.

- g. If Microsoft monitors for the activities described in Requests 8 (a)–(f) above, how does it 1) do so while still respecting the privacy of its users and 2) differentiate between legitimate communications – such as a business sending a mass email promotion – and nefarious ones?
 - h. If Microsoft identifies the content described in Requests 8 (a)–(f), what actions, if any, does Microsoft take to alert law enforcement and/or any of the individuals or entities that were impersonated or targeted? If Microsoft takes no action, please explain.
- 9. What steps does Microsoft take to prevent Personally Identifiable Information (PII) contained in illicit sources – such as from data breaches, hacks, or other leaks of sensitive and/or classified information from public or private sector sources – from being used as part of its training data?
 - a. Does Microsoft conduct regular training data audits to ensure that illicitly-sourced PII is not in its training data? If so, how often are such audits conducted?
 - b. Has Microsoft alerted government officials after discovering PII that was illegally sourced? If so, please describe the frequency of these alerts and the entities contacted.
- 10. Has Microsoft provided a way for an individual with information online to opt out of having their PII or other sensitive information included in the training content of its products?
- 11. For the users whose personal data is being used to train Microsoft’s models, is the company detecting and/or removing any PII or other sensitive information in users’ chat histories – including discussions of any sensitive financial information, medical history, etc., for any individual – from training material?
- 12. What steps does Microsoft take prior to a product release to ensure that its programs will not release PII? Have concerns over PII ever delayed Microsoft in releasing a product? If so, which products were delayed?
- 13. Can Microsoft detect if the programs it has released to the public share PII with other users, and if so, how? How does Microsoft monitor its programs in real time, if at all, to identify, prevent, and correct such incidents?
- 14. If a Microsoft product improperly shares PII, does the company inform the individuals whose data has been released?

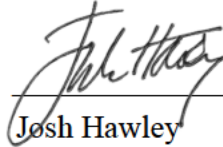
15. Please describe any cooperation or coordination between Microsoft and law enforcement agencies at the international, federal, state, and/or local levels to prevent or investigate the use of its products for fraudulent purposes.
16. Please describe any cooperation or coordination between Microsoft and government agencies that are responsible for combating fraud, including but not limited to the Federal Trade Commission.

Please provide your responses as soon as possible but in no event later than January 14, 2026. If you have any questions related to this request, please contact [REDACTED] of the Joint Economic Committee staff at [REDACTED]. Please send any official correspondence relating to this request to [REDACTED].

Sincerely,



Margaret Wood Hassan
Ranking Member
U.S. Congress Joint Economic
Committee



Josh Hawley
U.S. Senator

cc: David Schweikert
Chairman, Joint Economic Committee

Eric Schmitt
Vice Chairman, Joint Economic Committee

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mr. Sam Altman
Chief Executive Officer
OpenAI
1455 3rd Street
San Francisco, CA 94158

Dear Mr. Altman:

We write today to request information on OpenAI's efforts to implement sufficient protections to prevent scammers from misusing its technology. In recent years, global criminal networks have turned to AI to target more people with increasingly personalized and believable digital scams, contributing to a booming scam economy that is now a larger illicit industry than the global drug trade.¹ The FBI recorded \$16.6 billion in reported losses to suspected scams and other cybercrime in 2024 – up from \$3.5 billion in 2019, a more than 370 percent increase over the last five years.² The federal government has a responsibility to protect the American people from scams, but this effort requires an all-hands-on-deck approach across multiple industries. Given your public statements about the potential for AI to contribute to a “fraud crisis,” we hope we can work together productively on this important issue.³

With advancements in AI, scams will continue to grow in sophistication, frequency, and impact.⁴ In the early phases of a scam, criminals can use generative AI to quickly identify and

¹ *Online Scams May Already Be as Big a Scourge as Illegal Drugs*, The Economist (Feb. 6, 2025) (www.economist.com/briefing/2025/02/06/online-scams-may-already-be-as-big-a-scourge-as-illegal-drugs); The Economist Group, *Press Release: Fraud Factories, Cyber Criminals and Corruption: The Economist's New Podcast, "Scam Inc", Uncovers a New, Global, Underground Economy Worth More Than Illicit Drug Trade* (Feb. 6, 2025) (www.economistgroup.com/press-centre/the-economist/fraud-factories-cyber-criminals-and-corruption-the-economists-new-podcast-scam-inc).

² FBI, *Internet Crime Report* (2024) (www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf); FBI, *Internet Crime Report* (2019) (www.ic3.gov/AnnualReport/Reports/2019_ic3Report.pdf).

³ LIVE: OpenAI CEO Sam Altman Speaks with Fed's Michelle Bowman on Bank Capital Rules, Associated Press (July 22, 2025) (www.youtube.com/watch?v=tScbQiavmpA).

⁴ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

then collect details on their targets, enabling them to create tailor-made scams.⁵ Once armed with information like addresses, account numbers, and birthdates, bad actors can more realistically impersonate a victim's bank, a government office, or even a member of their own family in an attempt to gain control of accounts or induce fraudulent payments.⁶

Fraudsters can use AI models to craft and then send convincing email, text, and phone call scams at an industrial scale.⁷ As Deloitte's Center for Financial Services described, "fake content has never been easier to create – or harder to catch."⁸ In one growing type of scam, criminals have called victims using AI-generated voice clones of their loved ones and pretended to need money after an emergency.⁹ For instance, in June 2025, a New York man was sentenced to prison for his role in a high-tech, "elaborate grandparent scam" in which he stole around \$20,000 from three New Hampshire families after convincing them that their loved ones were in trouble.¹⁰ According to the *Union Leader*, "[v]ictims say the scam involved the use of artificial intelligence mimicking a loved one's voice to trick them into turning over money to bail that person out of jail."¹¹ Research has shown that these calls can now be fully automated with the

⁵ Fredrik Heiding, Bruce Schneier, and Arun Vishwanath, *AI Will Increase the Quantity — and Quality — of Phishing Scams*, Harvard Business Review (May 30, 2024) (hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams).

⁶ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁷ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/); U.S. PIRG Education Fund, *A Dark Side of AI: Text and Phone Scams Are Getting Worse* (Nov. 14, 2025) (pirg.org/edfund/articles/a-dark-side-of-ai-text-and-phone-scams-are-getting-worse/).

⁸ Deloitte's Center for Financial Services, *Generative AI is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking* (May 29, 2024) (www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html).

⁹ Federal Trade Commission, *Scammers Use AI to Enhance Their Family Emergency Schemes* (Mar. 20, 2023) (consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

¹⁰ *Bronx Man Gets 4 To 8 Years for Role in 'Grandparent Scam,'* New Hampshire Union Leader (Sept. 24, 2025) (www.unionleader.com/news/courts/bronx-man-gets-4-to-8-years-for-role-in-grandparent-scam/article_58a05fcd-5edd-4d76-8dae-8ccd3937dcac.html).

¹¹ *Id.*

use of interactive AI models capable of carrying on real-time conversations, allowing scammers to potentially reach many people at little cost – but massive harm to victims.¹² Bad actors are using this same voice-altering technology to imitate employees of government agencies, banks, utilities, and other offices.¹³

Impersonation scams are also executed over email, and AI has made fraudulent email messages “so good you simply cannot tell” that they are fake.¹⁴ All kinds of organizations have been impersonated in email phishing attacks, including federal agencies.¹⁵ The Department of Veterans Affairs (VA), for example, has warned veterans of emails that look like official communications and seek to extract benefits and personal information.¹⁶ The VA has also specifically spotlighted how scammers are using generative AI to target veterans in these fraudulent email campaigns and other scams.¹⁷ The Social Security Administration (SSA) Office of the Inspector General has also issued multiple alerts this year warning beneficiaries to look out for realistic emails that purport to be from SSA and try to lure users to a fraudulent website or trick them into downloading malicious software.¹⁸

¹² *Voice-Enabled AI Agents Can Automate Everything, Even Your Phone Scams*, The Register (Oct. 24, 2024) (www.theregister.com/2024/10/24/openai_realtime_api_phone_scam/).

¹³ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert).

¹⁴ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

¹⁵ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers); Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/).

¹⁶ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers).

¹⁷ Department of Veterans Affairs, *Protecting Yourself and Your Benefits from Cyber & Artificial Intelligence (AI) Threats* (Apr. 29, 2025) (department.va.gov/privacy/fact-sheet/protecting-yourself-and-your-benefits-from-cyber-artificial-intelligence-ai-threats/).

¹⁸ Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/); Office of the Inspector General, Social Security Administration, *Press Release: Beware of Scam Emails Asking to Download Statements* (Apr. 1, 2025) (oig.ssa.gov/scam-alerts/2025-04-01-beware-of-scam-emails-asking-to-download-statements/).

OpenAI rightly prohibits the use of its products to perpetrate fraud and scams.¹⁹ AI companies, however, have reportedly faced challenges in preventing the misuse of their technology.²⁰ For example, users have tricked AI programs into performing prohibited tasks simply through the “clever phrasing” of a prompt, according to a Fox News report, and *Scientific American* has similarly reported that researchers have “repeatedly demonstrated ways to get around” measures designed to prevent users from using AI programs in harmful ways.²¹ In addition, AI systems are often trained on sensitive data, with one recent study uncovering credit card and passport information in a widely used training set.²² Research has shown that “these systems can memorize the data and then expose it to other users as part of the outputs.”²³ In the hands of a fraudster, this sensitive information could be used to develop a highly personalized scam or to steal someone’s identity outright.²⁴ These urgent and complex challenges call for collaboration between the private and public sectors, and we want to work together to find solutions to protect Americans. Accordingly, we seek responses to the requests below to better understand OpenAI’s operations and commitment to preventing AI-driven scams.

The following questions pertain to all iterations of all generative AI services and models that OpenAI has released to the public, including those available only in limited or beta release.

1. Describe OpenAI’s overall strategy to prevent the misuse of its technologies for scams and fraud – two activities that are in violation of its usage policies.

¹⁹ OpenAI, Usage Policies (openai.com/policies/usage-policies/) (accessed Dec. 2, 2025).

²⁰ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, Scientific American (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²¹ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, Scientific American (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²² Rachel Hong et al., *A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset*, arXiv:2506.17185 [cs.CR] (June 20, 2025) (arxiv.org/html/2506.17185v1#bib).

²³ Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era*, Stanford University Human-Centered Artificial Intelligence (Feb. 2024) (hai.stanford.edu/assets/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf).

²⁴ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

- a. What alternative fraud-prevention strategies has your company considered or tested?
 - b. How much does the company invest to prevent scams and fraud?
 - c. How many people does OpenAI employ, either as employees of the company or outside consultants?
 - d. How many people does OpenAI employ, either as employees of the company or outside consultants, to work on preventing scams?
2. Does OpenAI monitor for user activity that violates the company's usage policies prohibiting the use of its models for scams and fraud?
 - a. Why or why not?
3. If OpenAI does track violations of its policies prohibiting the use of its services to scam or defraud, please provide annual data for each OpenAI model and version regarding:
 - a. The number of instances in which OpenAI's proactive detection tools identified content or inputs that violated policies prohibiting the use of its services to scam or defraud;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - b. The number of instances in which OpenAI was alerted to potential violations of its policies prohibiting the use of its services to scam or defraud by user reports or external notices;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - c. The number of individual accounts and individual users whose accounts OpenAI has deactivated and/or suspended for violating its policies against scams and fraud.
 - i. Please provide the number of these deactivations and suspensions that occurred following human review, if any. Please explain the level of review OpenAI requires prior to these actions, in general, and the rationale for these procedures.
 - d. The number of individual accounts and individual users against which OpenAI has taken action short of account suspensions or terminations, including through warnings, sharing restrictions, or ineligibility for inclusion in GPT Store or monetization, broken down by type of action;

- i. Please provide the number of actions that were taken following human review, if any. Please explain the level of review OpenAI requires prior to these actions, in general, and the rationale for these procedures.
 - e. A breakdown of the severity and nature of the activities that OpenAI determined were in violation of its policies against scams and fraud (i.e., the specific usage policy or policies violated);
 - i. Of these violations, please state the number OpenAI initially identified via proactive detection, user reports or external notices, and human review, respectively.
- 4. Please detail the steps OpenAI takes to authenticate users, specifically:
 - a. Please describe what, if any, know-your-customer information OpenAI collects, including but not limited to user names, emails, phone numbers, credit card information, IP addresses, and physical location. Please describe efforts by OpenAI, if any, to authenticate these user details.
 - b. Please detail the steps that the company takes to ensure that banned users are prohibited from creating new OpenAI accounts or operating additional existing accounts under a different name and/or with fraudulent credentials.
 - c. If OpenAI detects someone whom the platform banned for illegal activity attempting to sign up for a new account or otherwise operate another account, does OpenAI alert law enforcement? If so, under what circumstances does OpenAI make these alerts? If not, why not?
- 5. Does OpenAI attempt to determine if users and/or accounts that violate its policies regarding scams and fraud are performing these activities in coordination with other parties?
 - a. For example, does OpenAI look for evidence or patterns that suggest violators may be working as part of a broader criminal network or on behalf of a foreign government?
 - b. Similarly, does OpenAI try to determine if these user accounts or activities belong to or are otherwise tied to individuals the United States has designated as threats, including Specially Designated Nationals, other sanctioned individuals, individuals designated as foreign or domestic terrorists, or wanted criminals?
 - i. If so, how many such user accounts has OpenAI identified?
 - ii. If so, does OpenAI alert law enforcement to such instances? If so, under what circumstances does OpenAI make these alerts? If not, why not?

6. Please describe OpenAI policies and procedures regarding location-specific access restrictions, also known as “geofencing,” including with respect to restrictions designed to prevent the use of its products in known hotspots for scam activity.
7. How many outputs have OpenAI’s models generated for users that violated the company’s own usage policies with respect to scams and/or fraud, including but not limited to providing written information, producing visual content, or performing functions that were impermissible under the company’s rules?
 - a. What share of these model outputs were in response to user inputs that violated the company’s usage policies regarding scams and/or fraud?
 - b. What share of these model outputs were in response to user inputs considered acceptable under the company’s usage policies?
8. Please describe any measures OpenAI has in place to monitor for and/or prevent the user activities listed below that can be used to facilitate scams and/or fraud.
 - a. Recreating logos of an entity, including banks and government offices, without permission;
 - b. Drafting emails, text messages, videos, or voice messages, including ones that utilize AI-generated voice technology, that suggest that the user is a trusted entity like a bank or government agency;
 - c. Drafting or preparing recipient lists for emails, text messages, videos, or voice messages – particularly to a high volume of recipients – that ask the recipient to make a payment or provide sensitive personal information, such as bank account numbers, identifying information, or login credentials, or direct the recipient to a website designed to collect such a payment or personal information;
 - d. Drafting or preparing recipient lists for emails or text messages that contain malware, viruses, or other malicious content designed to take over or otherwise corrupt the recipient’s accounts, software, or hardware;
 - e. Inputs designed to facilitate the theft of passwords or login credentials; and
 - f. Inputs designed to bypass biometric account verification measures such as facial recognition.
 - g. If OpenAI monitors for the activities described in Requests 8 (a)–(f) above, how does it 1) do so while still respecting the privacy of its users and 2) differentiate between legitimate communications – such as a business sending a mass email promotion – and nefarious ones?

- h. If OpenAI identifies the content described in Requests 8 (a)–(f), what actions, if any, does OpenAI take to alert law enforcement and/or any of the individuals or entities that were impersonated or targeted? If OpenAI takes no action, please explain.
- 9. What steps does OpenAI take to prevent Personally Identifiable Information (PII) contained in illicit sources – such as from data breaches, hacks, or other leaks of sensitive and/or classified information from public or private sector sources – from being used as part of its training data?
 - a. Does OpenAI conduct regular training data audits to ensure that illicitly-sourced PII is not in its training data? If so, how often are such audits conducted?
 - b. Has OpenAI alerted government officials after discovering PII that was illegally sourced? If so, please describe the frequency of these alerts and the entities contacted.
- 10. Has OpenAI provided a way for an individual with information online to opt out of having their PII or other sensitive information included in the training content of its products?
- 11. For the users whose personal data is being used to train OpenAI’s models, is the company detecting and/or removing any PII or other sensitive information in users’ chat histories – including discussions of any sensitive financial information, medical history, etc., for any individual – from training material?
- 12. What steps does OpenAI take prior to a product release to ensure that its programs will not release PII? Have concerns over PII ever delayed OpenAI in releasing a product? If so, which products were delayed?
- 13. Can OpenAI detect if the programs it has released to the public share PII with other users, and if so, how? How does OpenAI monitor its programs in real time, if at all, to identify, prevent, and correct such incidents?
- 14. If an OpenAI product improperly shares PII, does the company inform the individuals whose data has been released?
- 15. Please describe any cooperation or coordination between OpenAI and law enforcement agencies at the international, federal, state, and/or local levels to prevent or investigate the use of its products for fraudulent purposes.

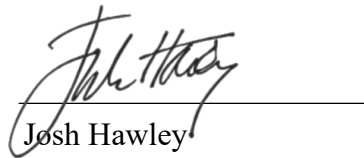
16. Please describe any cooperation or coordination between OpenAI and government agencies that are responsible for combating fraud, including but not limited to the Federal Trade Commission.

Please provide your responses as soon as possible but in no event later than January 14, 2026. If you have any questions related to this request, please contact [REDACTED] of the Joint Economic Committee staff at [REDACTED]. Please send any official correspondence relating to this request to [REDACTED].

Sincerely,



Margaret Wood Hassan
Ranking Member
U.S. Congress Joint Economic
Committee



Josh Hawley
U.S. Senator

cc: David Schweikert
Chairman, Joint Economic Committee

Eric Schmitt
Vice Chairman, Joint Economic Committee

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mr. Aravind Srinivas
Chief Executive Officer
Perplexity
115 Sansome Street
Suite 900
San Francisco, CA 94104

Dear Mr. Srinivas:

We write today to request information on Perplexity's efforts to implement sufficient protections to prevent scammers from misusing its technology. In recent years, global criminal networks have turned to AI to target more people with increasingly personalized and believable digital scams, contributing to a booming scam economy that is now a larger illicit industry than the global drug trade.¹ The FBI recorded \$16.6 billion in reported losses to suspected scams and other cybercrime in 2024 – up from \$3.5 billion in 2019, a more than 370 percent increase over the last five years.² The federal government has a responsibility to protect the American people from scams, but this effort requires an all-hands-on-deck approach across multiple industries. We hope we can work together productively on this important issue.

With advancements in AI, scams will continue to grow in sophistication, frequency, and impact.³ In the early phases of a scam, criminals can use generative AI to quickly identify and

¹ *Online Scams May Already Be as Big a Scourge as Illegal Drugs*, The Economist (Feb. 6, 2025) (www.economist.com/briefing/2025/02/06/online-scams-may-already-be-as-big-a-scourge-as-illegal-drugs); The Economist Group, *Press Release: Fraud Factories, Cyber Criminals and Corruption: The Economist's New Podcast, "Scam Inc", Uncovers a New, Global, Underground Economy Worth More Than Illicit Drug Trade* (Feb. 6, 2025) (www.economistgroup.com/press-centre/the-economist/fraud-factories-cyber-criminals-and-corruption-the-economists-new-podcast-scam-inc).

² FBI, *Internet Crime Report* (2024) (www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf); FBI, *Internet Crime Report* (2019) (www.ic3.gov/AnnualReport/Reports/2019_ic3Report.pdf).

³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

then collect details on their targets, enabling them to create tailor-made scams.⁴ Once armed with information like addresses, account numbers, and birthdates, bad actors can more realistically impersonate a victim's bank, a government office, or even a member of their own family in an attempt to gain control of accounts or induce fraudulent payments.⁵

Fraudsters can use AI models to craft and then send convincing email, text, and phone call scams at an industrial scale.⁶ As Deloitte's Center for Financial Services described, "fake content has never been easier to create – or harder to catch."⁷ In one growing type of scam, criminals have called victims using AI-generated voice clones of their loved ones and pretended to need money after an emergency.⁸ For instance, in June 2025, a New York man was sentenced to prison for his role in a high-tech, "elaborate grandparent scam" in which he stole around \$20,000 from three New Hampshire families after convincing them that their loved ones were in trouble.⁹ According to the *Union Leader*, "[v]ictims say the scam involved the use of artificial intelligence mimicking a loved one's voice to trick them into turning over money to bail that person out of jail."¹⁰ Research has shown that these calls can now be fully automated with the

⁴ Fredrik Heiding, Bruce Schneier, and Arun Vishwanath, *AI Will Increase the Quantity — and Quality — of Phishing Scams*, Harvard Business Review (May 30, 2024) (hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams).

⁵ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁶ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/); U.S. PIRG Education Fund, *A Dark Side of AI: Text and Phone Scams Are Getting Worse* (Nov. 14, 2025) (pirg.org/edfund/articles/a-dark-side-of-ai-text-and-phone-scams-are-getting-worse/).

⁷ Deloitte's Center for Financial Services, *Generative AI is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking* (May 29, 2024) (www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html).

⁸ Federal Trade Commission, *Scammers Use AI to Enhance Their Family Emergency Schemes* (Mar. 20, 2023) (consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁹ *Bronx Man Gets 4 To 8 Years for Role in 'Grandparent Scam,'* New Hampshire Union Leader (Sept. 24, 2025) (www.unionleader.com/news/courts/bronx-man-gets-4-to-8-years-for-role-in-grandparent-scam/article_58a05fcd-5edd-4d76-8dae-8ccd3937dcac.html).

¹⁰ *Id.*

use of interactive AI models capable of carrying on real-time conversations, allowing scammers to potentially reach many people at little cost – but massive harm to victims.¹¹ Bad actors are using this same voice-altering technology to imitate employees of government agencies, banks, utilities, and other offices.¹²

Impersonation scams are also executed over email, and AI has made fraudulent email messages “so good you simply cannot tell” that they are fake.¹³ All kinds of organizations have been impersonated in email phishing attacks, including federal agencies.¹⁴ The Department of Veterans Affairs (VA), for example, has warned veterans of emails that look like official communications and seek to extract benefits and personal information.¹⁵ The VA has also specifically spotlighted how scammers are using generative AI to target veterans in these fraudulent email campaigns and other scams.¹⁶ The Social Security Administration (SSA) Office of the Inspector General has also issued multiple alerts this year warning beneficiaries to look out for realistic emails that purport to be from SSA and try to lure users to a fraudulent website or trick them into downloading malicious software.¹⁷

¹¹ *Voice-Enabled AI Agents Can Automate Everything, Even Your Phone Scams*, The Register (Oct. 24, 2024) (www.theregister.com/2024/10/24/openai_realtime_api_phone_scam/).

¹² *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert).

¹³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

¹⁴ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers); Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/).

¹⁵ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers).

¹⁶ Department of Veterans Affairs, *Protecting Yourself and Your Benefits from Cyber & Artificial Intelligence (AI) Threats* (Apr. 29, 2025) (department.va.gov/privacy/fact-sheet/protecting-yourself-and-your-benefits-from-cyber-artificial-intelligence-ai-threats/).

¹⁷ Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/); Office of the Inspector General, Social Security Administration, *Press Release: Beware of Scam Emails Asking to Download Statements* (Apr. 1, 2025) (oig.ssa.gov/scam-alerts/2025-04-01-beware-of-scam-emails-asking-to-download-statements/).

Perplexity rightly prohibits the use of its products for fraudulent and unlawful activity, including specific actions commonly associated with scams, such as “hacking, phishing, identity theft, or creation or distribution of malware.”¹⁸ AI companies, however, have reportedly faced challenges in preventing the misuse of their technology.¹⁹ For example, users have tricked AI programs into performing prohibited tasks simply through the “clever phrasing” of a prompt, according to a Fox News report, and *Scientific American* has similarly reported that researchers have “repeatedly demonstrated ways to get around” measures designed to prevent users from using AI programs in harmful ways.²⁰ In addition, AI systems are often trained on sensitive data, with one recent study uncovering credit card and passport information in a widely used training set.²¹ Research has shown that “these systems can memorize the data and then expose it to other users as part of the outputs.”²² In the hands of a fraudster, this sensitive information could be used to develop a highly personalized scam or to steal someone’s identity outright.²³ These urgent and complex challenges call for collaboration between the private and public sectors, and we want to work together to find solutions to protect Americans. Accordingly, we seek responses to the requests below to better understand Perplexity’s operations and commitment to preventing AI-driven scams.

The following questions pertain to all iterations of all generative AI services and models that Perplexity has released to the public, including those available only in limited or beta release.

¹⁸ Perplexity, Acceptable Use Policy (www.perplexity.ai/hub/legal/aup) (accessed Dec. 3, 2025).

¹⁹ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²⁰ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, *Scientific American* (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²¹ Rachel Hong et al., *A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset*, arXiv:2506.17185 [cs.CR] (June 20, 2025) (arxiv.org/html/2506.17185v1#bib).

²² Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era*, Stanford University Human-Centered Artificial Intelligence (Feb. 2024) (hai.stanford.edu/assets/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf).

²³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, *Forbes* (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

1. Describe Perplexity's overall strategy to prevent the misuse of its technologies for scams and fraud – two activities that are in violation of its usage policies.
 - a. What alternative fraud-prevention strategies has your company considered or tested?
 - b. How much does the company invest to prevent scams and fraud?
 - c. How many people does Perplexity employ, either as employees of the company or outside consultants?
 - d. How many people does Perplexity employ, either as employees of the company or outside consultants, to work on preventing scams?
2. Does Perplexity monitor for user activity that violates the company's usage policies prohibiting the use of its models for scams and fraud?
 - a. Why or why not?
3. If Perplexity does track violations of its policies prohibiting the use of its services to scam or defraud, please provide annual data for each Perplexity model and version regarding:
 - a. The number of instances in which any proactive detection tools used by Perplexity identified content or inputs that violated policies prohibiting the use of its services to scam or defraud;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - b. The number of instances in which Perplexity was alerted to potential violations of its policies prohibiting the use of its services to scam or defraud by user reports or external notices;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - c. The number of individual accounts and individual users whose accounts Perplexity has deactivated and/or suspended for violating its policies against scams and fraud.
 - i. Please provide the number of these deactivations and suspensions that occurred following human review, if any. Please explain the level of review Perplexity requires prior to these actions, in general, and the rationale for these procedures.

- d. The number of individual accounts and individual users against which Perplexity has taken action short of account suspensions or terminations, including through warnings or account restrictions, broken down by type of action;
 - i. Please provide the number of actions that were taken following human review, if any. Please explain the level of review Perplexity requires prior to these actions, in general, and the rationale for these procedures.
 - e. A breakdown of the severity and nature of the activities that Perplexity determined were in violation of its policies against scams and fraud (i.e., the specific usage policy or policies violated);
 - i. Of these violations, please state the number Perplexity initially identified via proactive detection, user reports or external notices, and human review, respectively.
4. Please detail the steps Perplexity takes to authenticate users, specifically:
- a. Please describe what, if any, know-your-customer information Perplexity collects, including but not limited to user names, emails, phone numbers, credit card information, IP addresses, and physical location. Please describe efforts by Perplexity, if any, to authenticate these user details.
 - b. Please detail the steps that the company takes to ensure that banned users are prohibited from creating new Perplexity accounts or operating additional existing accounts under a different name and/or with fraudulent credentials.
 - c. If Perplexity detects someone whom the platform banned for illegal activity attempting to sign up for a new account or otherwise operate another account, does Perplexity alert law enforcement? If so, under what circumstances does Perplexity make these alerts? If not, why not?
5. Does Perplexity attempt to determine if users and/or accounts that violate its policies regarding scams and fraud are performing these activities in coordination with other parties?
- a. For example, does Perplexity look for evidence or patterns that suggest violators may be working as part of a broader criminal network or on behalf of a foreign government?
 - b. Similarly, does Perplexity try to determine if these user accounts or activities belong to or are otherwise tied to individuals the United States has designated as threats, including Specially Designated Nationals, other sanctioned individuals, individuals designated as foreign or domestic terrorists, or wanted criminals?

- i. If so, how many such user accounts has Perplexity identified?
 - ii. If so, does Perplexity alert law enforcement to such instances? If so, under what circumstances does Perplexity make these alerts? If not, why not?
6. Please describe Perplexity's policies and procedures regarding location-specific access restrictions, also known as "geofencing," including with respect to restrictions designed to prevent the use of its products in known hotspots for scam activity.
7. How many outputs have Perplexity's models generated for users that violated the company's own usage policies with respect to scams and/or fraud, including but not limited to providing written information, producing visual content, or performing functions that were impermissible under the company's rules?
 - a. What share of these model outputs were in response to user inputs that violated the company's usage policies regarding scams and/or fraud?
 - b. What share of these model outputs were in response to user inputs considered acceptable under the company's usage policies?
8. Please describe any measures Perplexity has in place to monitor for and/or prevent the user activities listed below that can be used to facilitate scams and/or fraud.
 - a. Recreating logos of an entity, including banks and government offices, without permission;
 - b. Drafting emails, text messages, videos, or voice messages, including ones that utilize AI-generated voice technology, that suggest that the user is a trusted entity like a bank or government agency;
 - c. Drafting or preparing recipient lists for emails, text messages, videos, or voice messages – particularly to a high volume of recipients – that ask the recipient to make a payment or provide sensitive personal information, such as bank account numbers, identifying information, or login credentials, or direct the recipient to a website designed to collect such a payment or personal information;
 - d. Drafting or preparing recipient lists for emails or text messages that contain malware, viruses, or other malicious content designed to take over or otherwise corrupt the recipient's accounts, software, or hardware;
 - e. Inputs designed to facilitate the theft of passwords or login credentials; and
 - f. Inputs designed to bypass biometric account verification measures such as facial recognition.

- g. If Perplexity monitors for the activities described in Requests 8 (a)–(f) above, how does it 1) do so while still respecting the privacy of its users and 2) differentiate between legitimate communications – such as a business sending a mass email promotion – and nefarious ones?
 - h. If Perplexity identifies the content described in Requests 8 (a)–(f), what actions, if any, does Perplexity take to alert law enforcement and/or any of the individuals or entities that were impersonated or targeted? If Perplexity takes no action, please explain.
- 9. What steps does Perplexity take to prevent Personally Identifiable Information (PII) contained in illicit sources – such as from data breaches, hacks, or other leaks of sensitive and/or classified information from public or private sector sources – from being used as part of its training data?
 - a. Does Perplexity conduct regular training data audits to ensure that illicitly-sourced PII is not in its training data? If so, how often are such audits conducted?
 - b. Has Perplexity alerted government officials after discovering PII that was illegally sourced? If so, please describe the frequency of these alerts and the entities contacted.
- 10. Has Perplexity provided a way for an individual with information online to opt out of having their PII or other sensitive information included in the training content of its products?
- 11. For the users whose personal data is being used to train Perplexity’s models, is the company detecting and/or removing any PII or other sensitive information in users’ chat and search histories – including discussions of any sensitive financial information, medical history, etc., for any individual – from training material?
- 12. What steps does Perplexity take prior to a product release to ensure that its programs will not release PII? Have concerns over PII ever delayed Perplexity in releasing a product? If so, which products were delayed?
- 13. Can Perplexity detect if the programs it has released to the public share PII with other users, and if so, how? How does Perplexity monitor its programs in real time, if at all, to identify, prevent, and correct such incidents?
- 14. If a Perplexity product improperly shares PII, does the company inform the individuals whose data has been released?

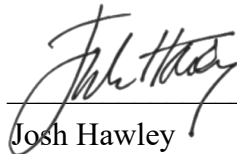
15. Please describe any cooperation or coordination between Perplexity and law enforcement agencies at the international, federal, state, and/or local levels to prevent or investigate the use of its products for fraudulent purposes.
16. Please describe any cooperation or coordination between Perplexity and government agencies that are responsible for combating fraud, including but not limited to the Federal Trade Commission.

Please provide your responses as soon as possible but in no event later than January 14, 2026. If you have any questions related to this request, please contact [REDACTED] of the Joint Economic Committee staff at [REDACTED]. Please send any official correspondence relating to this request to [REDACTED].

Sincerely,



Margaret Wood Hassan
Ranking Member
U.S. Congress Joint Economic
Committee



Josh Hawley
U.S. Senator

cc: David Schweikert
Chairman, Joint Economic Committee

Eric Schmitt
Vice Chairman, Joint Economic Committee

United States Senate

WASHINGTON, DC 20510

December 11, 2025

Mr. Elon Musk
Chief Executive Officer
xAI
1450 Page Mill Road
Palo Alto, CA 94304

Dear Mr. Musk:

We write today to request information on xAI's efforts to implement sufficient protections to prevent scammers from misusing its technology. In recent years, global criminal networks have turned to AI to target more people with increasingly personalized and believable digital scams, contributing to a booming scam economy that is now a larger illicit industry than the global drug trade.¹ The FBI recorded \$16.6 billion in reported losses to suspected scams and other cybercrime in 2024 – up from \$3.5 billion in 2019, a more than 370 percent increase over the last five years.² The federal government has a responsibility to protect the American people from scams, but this effort requires an all-hands-on-deck approach across multiple industries. We hope we can work together productively on this important issue.

With advancements in AI, scams will continue to grow in sophistication, frequency, and impact.³ In the early phases of a scam, criminals can use generative AI to quickly identify and

¹ *Online Scams May Already Be as Big a Scourge as Illegal Drugs*, The Economist (Feb. 6, 2025) (www.economist.com/briefing/2025/02/06/online-scams-may-already-be-as-big-a-scourge-as-illegal-drugs); The Economist Group, *Press Release: Fraud Factories, Cyber Criminals and Corruption: The Economist's New Podcast, "Scam Inc", Uncovers a New, Global, Underground Economy Worth More Than Illicit Drug Trade* (Feb. 6, 2025) (www.economistgroup.com/press-centre/the-economist/fraud-factories-cyber-criminals-and-corruption-the-economists-new-podcast-scam-inc).

² FBI, *Internet Crime Report* (2024) (www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf); FBI, *Internet Crime Report* (2019) (www.ic3.gov/AnnualReport/Reports/2019_ic3Report.pdf).

³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

then collect details on their targets, enabling them to create tailor-made scams.⁴ Once armed with information like addresses, account numbers, and birthdates, bad actors can more realistically impersonate a victim's bank, a government office, or even a member of their own family in an attempt to gain control of accounts or induce fraudulent payments.⁵

Fraudsters can use AI models to craft and then send convincing email, text, and phone call scams at an industrial scale.⁶ As Deloitte's Center for Financial Services described, "fake content has never been easier to create – or harder to catch."⁷ In one growing type of scam, criminals have called victims using AI-generated voice clones of their loved ones and pretended to need money after an emergency.⁸ For instance, in June 2025, a New York man was sentenced to prison for his role in a high-tech, "elaborate grandparent scam" in which he stole around \$20,000 from three New Hampshire families after convincing them that their loved ones were in trouble.⁹ According to the *Union Leader*, "[v]ictims say the scam involved the use of artificial intelligence mimicking a loved one's voice to trick them into turning over money to bail that person out of jail."¹⁰ Research has shown that these calls can now be fully automated with the

⁴ Fredrik Heiding, Bruce Schneier, and Arun Vishwanath, *AI Will Increase the Quantity — and Quality — of Phishing Scams*, Harvard Business Review (May 30, 2024) (hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams).

⁵ *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁶ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/); U.S. PIRG Education Fund, *A Dark Side of AI: Text and Phone Scams Are Getting Worse* (Nov. 14, 2025) (pirg.org/edfund/articles/a-dark-side-of-ai-text-and-phone-scams-are-getting-worse/).

⁷ Deloitte's Center for Financial Services, *Generative AI is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking* (May 29, 2024) (www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html).

⁸ Federal Trade Commission, *Scammers Use AI to Enhance Their Family Emergency Schemes* (Mar. 20, 2023) (consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes); *AI Phone Scam Mimicking Your Voice Can Now Be Used to Scam Family, Friends, Experts Warn*, ABC7 (May 20, 2024) (abc7chicago.com/post/ai-phone-scam-calls-mimicking-voice-scam-family/14847406/).

⁹ *Bronx Man Gets 4 To 8 Years for Role in 'Grandparent Scam,'* New Hampshire Union Leader (Sept. 24, 2025) (www.unionleader.com/news/courts/bronx-man-gets-4-to-8-years-for-role-in-grandparent-scam/article_58a05fcd-5edd-4d76-8dae-8ccd3937dcac.html).

¹⁰ *Id.*

use of interactive AI models capable of carrying on real-time conversations, allowing scammers to potentially reach many people at little cost – but massive harm to victims.¹¹ Bad actors are using this same voice-altering technology to imitate employees of government agencies, banks, utilities, and other offices.¹²

Impersonation scams are also executed over email, and AI has made fraudulent email messages “so good you simply cannot tell” that they are fake.¹³ All kinds of organizations have been impersonated in email phishing attacks, including federal agencies.¹⁴ The Department of Veterans Affairs (VA), for example, has warned veterans of emails that look like official communications and seek to extract benefits and personal information.¹⁵ The VA has also specifically spotlighted how scammers are using generative AI to target veterans in these fraudulent email campaigns and other scams.¹⁶ The Social Security Administration (SSA) Office of the Inspector General has also issued multiple alerts this year warning beneficiaries to look out for realistic emails that purport to be from SSA and try to lure users to a fraudulent website or trick them into downloading malicious software.¹⁷

¹¹ *Voice-Enabled AI Agents Can Automate Everything, Even Your Phone Scams*, The Register (Oct. 24, 2024) (www.theregister.com/2024/10/24/openai_realtime_api_phone_scam/).

¹² *FCC Names Its First-Ever AI Scammer in Threat Alert*, Fox News (June 5, 2024) (www.foxnews.com/tech/fcc-names-its-first-ever-ai-scammer-threat-alert).

¹³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

¹⁴ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers); Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/).

¹⁵ Department of Veterans Affairs, *How to Protect Your Identity and Your VA Benefits from Scammers* (Oct. 8, 2024) (www.va.gov/resources/how-to-protect-your-identity-and-your-va-benefits-from-scammers).

¹⁶ Department of Veterans Affairs, *Protecting Yourself and Your Benefits from Cyber & Artificial Intelligence (AI) Threats* (Apr. 29, 2025) (department.va.gov/privacy/fact-sheet/protecting-yourself-and-your-benefits-from-cyber-artificial-intelligence-ai-threats/).

¹⁷ Office of the Inspector General, Social Security Administration, *Press Release: It's a New Scam: "The Security Update Tool"* (July 7, 2025) (oig.ssa.gov/scam-alerts/2025-07-17-it%E2%80%99s-a-new-scam-%E2%80%9Cthe-security-update-tool%E2%80%9D/); Office of the Inspector General, Social Security Administration, *Press Release: Beware of Scam Emails Asking to Download Statements* (Apr. 1, 2025) (oig.ssa.gov/scam-alerts/2025-04-01-beware-of-scam-emails-asking-to-download-statements/).

xAI rightly prohibits the use of its products for “hacking, defrauding, ... scamming, spamming, or phishing.”¹⁸ AI companies, however, have reportedly faced challenges in preventing the misuse of their technology.¹⁹ For example, users have tricked AI programs into performing prohibited tasks simply through the “clever phrasing” of a prompt, according to a Fox News report, and *Scientific American* has similarly reported that researchers have “repeatedly demonstrated ways to get around” measures designed to prevent users from using AI programs in harmful ways.²⁰ In addition, AI systems are often trained on sensitive data, with one recent study uncovering credit card and passport information in a widely used training set.²¹ Research has shown that “these systems can memorize the data and then expose it to other users as part of the outputs.”²² In the hands of a fraudster, this sensitive information could be used to develop a highly personalized scam or to steal someone’s identity outright.²³ These urgent and complex challenges call for collaboration between the private and public sectors, and we want to work together to find solutions to protect Americans. Accordingly, we seek responses to the requests below to better understand xAI’s operations and commitment to preventing AI-driven scams.

The following questions pertain to all iterations of all generative AI services and models that xAI has released to the public, including those available only in limited or beta release.

¹⁸ xAI, Terms of Service – Consumer (x.ai/legal/terms-of-service/) (accessed Dec. 3, 2025).

¹⁹ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, Scientific American (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²⁰ *Scammers Can Exploit Your Data From Just 1 ChatGPT Search*, Fox News (May 28, 2025) (www.foxnews.com/tech/scammers-can-exploit-your-data-from-just-1-chatgpt-search); *Your Personal Information is Probably Being Used to Train Generative AI Models*, Scientific American (Oct. 19, 2023) (www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/).

²¹ Rachel Hong et al., *A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset*, arXiv:2506.17185 [cs.CR] (June 20, 2025) (arxiv.org/html/2506.17185v1#bib).

²² Jennifer King and Caroline Meinhardt, *Rethinking Privacy in the AI Era*, Stanford University Human-Centered Artificial Intelligence (Feb. 2024) (hai.stanford.edu/assets/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf).

²³ *New Gmail, Outlook, Apple Mail Warning — AI Nightmare Is Coming True In 2025*, Forbes (Jan. 3, 2025) (www.forbes.com/sites/zakdoffman/2025/01/03/new-gmail-outlook-apple-mail-warning-2025-hacking-nightmare-is-coming-true/).

1. Describe xAI's overall strategy to prevent the misuse of its technologies for scams and fraud – two activities that are in violation of its usage policies.
 - a. What alternative fraud-prevention strategies has your company considered or tested?
 - b. How much does the company invest to prevent scams and fraud?
 - c. How many people does xAI employ, either as employees of the company or outside consultants?
 - d. How many people does xAI employ, either as employees of the company or outside consultants, to work on preventing scams?
2. Does xAI monitor for user activity that violates the company's usage policies prohibiting the use of its models for scams and fraud?
 - a. Why or why not?
3. If xAI does track violations of its policies prohibiting the use of its services to scam or defraud, please provide annual data for each xAI model and version regarding:
 - a. The number of instances in which any proactive detection tools used by xAI identified content or inputs that violated policies prohibiting the use of its services to scam or defraud;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - b. The number of instances in which xAI was alerted to potential violations of its policies prohibiting the use of its services to scam or defraud by user reports or external notices;
 - i. How many violations initially detected in this manner were flagged for human review, and under what circumstances?
 - c. The number of individual accounts and individual users whose accounts xAI has deactivated and/or suspended for violating its policies against scams and fraud.
 - i. Please provide the number of these deactivations and suspensions that occurred following human review, if any. Please explain the level of review xAI requires prior to these actions, in general, and the rationale for these procedures.
 - d. The number of individual accounts and individual users against which xAI has taken action short of account suspensions or terminations, including through warnings or account restrictions, broken down by type of action;

- i. Please provide the number of actions that were taken following human review, if any. Please explain the level of review xAI requires prior to these actions, in general, and the rationale for these procedures.
 - e. A breakdown of the severity and nature of the activities that xAI determined were in violation of its policies against scams and fraud (i.e., the specific usage policy or policies violated);
 - i. Of these violations, please state the number xAI initially identified via proactive detection, user reports or external notices, and human review, respectively.
- 4. Please detail the steps xAI takes to authenticate users, specifically:
 - a. Please describe what, if any, know-your-customer information xAI collects, including but not limited to user names, emails, phone numbers, credit card information, IP addresses, and physical location. Please describe efforts by xAI, if any, to authenticate these user details.
 - b. Please detail the steps that the company takes to ensure that banned users are prohibited from creating new xAI accounts or operating additional existing accounts under a different name and/or with fraudulent credentials.
 - c. If xAI detects someone whom the platform banned for illegal activity attempting to sign up for a new account or otherwise operate another account, does xAI alert law enforcement? If so, under what circumstances does xAI make these alerts? If not, why not?
- 5. Does xAI attempt to determine if users and/or accounts that violate its policies regarding scams and fraud are performing these activities in coordination with other parties?
 - a. For example, does xAI look for evidence or patterns that suggest violators may be working as part of a broader criminal network or on behalf of a foreign government?
 - b. Similarly, does xAI try to determine if these user accounts or activities belong to or are otherwise tied to individuals the United States has designated as threats, including Specially Designated Nationals, other sanctioned individuals, individuals designated as foreign or domestic terrorists, or wanted criminals?
 - i. If so, how many such user accounts has xAI identified?
 - ii. If so, does xAI alert law enforcement to such instances? If so, under what circumstances does xAI make these alerts? If not, why not?

6. Please describe xAI's policies and procedures regarding location-specific access restrictions, also known as "geofencing," including with respect to restrictions designed to prevent the use of its products in known hotspots for scam activity.
7. How many outputs have xAI's models generated for users that violated the company's own usage policies with respect to scams and/or fraud, including but not limited to providing written information, producing visual content, or performing functions that were impermissible under the company's rules?
 - a. What share of these model outputs were in response to user inputs that violated the company's usage policies regarding scams and/or fraud?
 - b. What share of these model outputs were in response to user inputs considered acceptable under the company's usage policies?
8. Please describe any measures xAI has in place to monitor for and/or prevent the user activities listed below that can be used to facilitate scams and/or fraud.
 - a. Recreating logos of an entity, including banks and government offices, without permission;
 - b. Drafting emails, text messages, videos, or voice messages, including ones that utilize AI-generated voice technology, that suggest that the user is a trusted entity like a bank or government agency;
 - c. Drafting or preparing recipient lists for emails, text messages, videos, or voice messages – particularly to a high volume of recipients – that ask the recipient to make a payment or provide sensitive personal information, such as bank account numbers, identifying information, or login credentials, or direct the recipient to a website designed to collect such a payment or personal information;
 - d. Drafting or preparing recipient lists for emails or text messages that contain malware, viruses, or other malicious content designed to take over or otherwise corrupt the recipient's accounts, software, or hardware;
 - e. Inputs designed to facilitate the theft of passwords or login credentials; and
 - f. Inputs designed to bypass biometric account verification measures such as facial recognition.
 - g. If xAI monitors for the activities described in Requests 8 (a)–(f) above, how does it 1) do so while still respecting the privacy of its users and 2) differentiate between legitimate communications – such as a business sending a mass email promotion – and nefarious ones?

- h. If xAI identifies the content described in Requests 8 (a)–(f), what actions, if any, does xAI take to alert law enforcement and/or any of the individuals or entities that were impersonated or targeted? If xAI takes no action, please explain.
9. What steps does xAI take to prevent Personally Identifiable Information (PII) contained in illicit sources – such as from data breaches, hacks, or other leaks of sensitive and/or classified information from public or private sector sources – from being used as part of its training data?
 - a. Does xAI conduct regular training data audits to ensure that illicitly-sourced PII is not in its training data? If so, how often are such audits conducted?
 - b. Has xAI alerted government officials after discovering PII that was illegally sourced? If so, please describe the frequency of these alerts and the entities contacted.
10. Has xAI provided a way for an individual with information online to opt out of having their PII or other sensitive information included in the training content of its products?
11. For the users whose personal data is being used to train xAI's models, is the company detecting and/or removing any PII or other sensitive information in users' chat histories – including discussions of any sensitive financial information, medical history, etc., for any individual – from training material?
12. What steps does xAI take prior to a product release to ensure that its programs will not release PII? Have concerns over PII ever delayed xAI in releasing a product? If so, which products were delayed?
13. Can xAI detect if the programs it has released to the public share PII with other users, and if so, how? How does xAI monitor its programs in real time, if at all, to identify, prevent, and correct such incidents?
14. If an xAI product improperly shares PII, does the company inform the individuals whose data has been released?
15. Please describe any cooperation or coordination between xAI and law enforcement agencies at the international, federal, state, and/or local levels to prevent or investigate the use of its products for fraudulent purposes.
16. Please describe any cooperation or coordination between xAI and government agencies that are responsible for combating fraud, including but not limited to the Federal Trade Commission.

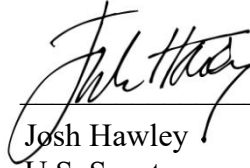
Mr. Elon Musk
December 11, 2025
Page 9

Please provide your responses as soon as possible but in no event later than January 14, 2026. If you have any questions related to this request, please contact [REDACTED] of the Joint Economic Committee staff at [REDACTED]. Please send any official correspondence relating to this request to [REDACTED].

Sincerely,



Margaret Wood Hassan
Ranking Member
U.S. Congress Joint Economic
Committee



Josh Hawley
U.S. Senator

cc: David Schweikert
Chairman, Joint Economic Committee

Eric Schmitt
Vice Chairman, Joint Economic Committee