United States Senate

June 22, 2023

The Honorable Gene L. Dodaro
Comptroller General of the United States
U.S. Government Accountability Office
441 G Street, Northwest
Washington, D.C. 20548

Dear Comptroller General Dodaro,

We write to ask the Government Accountability Office (GAO) to conduct a detailed technology assessment of the potential harms of generative artificial intelligence (AI) and how to mitigate them. The rollout of such popular chatbots such as ChatGPT — which respond to a user's inputs with natural language — and Midjourney — which generates images — has focused global attention on generative AI. Although generative AI holds the promise of many benefits, it is already causing significant harm.[1] In order to draw the maximum benefits from advances in AI, we must carefully study and understand its costs. Congress urgently requires the non-partisan, technical expertise that GAO is well placed to deliver.

We are early in the evolution of generative AI, but it promises tangible benefits to society if properly managed. Generative AI is already playing a role in the arts, sciences, law, business, and engineering. The technology can be used to speed up software development, and automate a variety of repetitive tasks.[2] It can accelerate scientific research,[3] and assist in the safe development of technologies such as autonomous vehicles.[4] Its responsible use in the federal government could improve the quality of government services for all Americans.[5] It is even preserving and spreading indigenous languages that are threatened.[6]

At the same time, it has already become apparent that generative AI is a double-edged sword, carrying with it a broad range of serious harms. Scammers have begun using generative AI for

---

[1] Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, Ass'n. for Comput. Mach. (Mar. 1, 2021), https://dl.acm.org/doi/10.1145/3442188.3445922; Electronic Privacy Information Center, *Generating Harms: Generative AI's Impact & Paths Forward*, (May 23, 2023), https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf.
[2] Government Accountability Office, *Science & Tech Spotlight: Generative AI*, (June 13, 2023), https://www.gao.gov/products/gao-23-106782.
[3] Megan Craig, *Using Generative AI to Accelerate the Drug Development Process.* Medical Life Sciences News (May 30, 2023), https://www.news-medical.net/news/20230530/Using-generative-AI-to-accelerate-the-drug-development-process.aspx.
[4] Sascha Dieh & Andrea Ketzer, *Harnessing the Power of Generative AI for Automotive Technologies on AWS*, AWS Blog (June 5, 2023), https://aws.amazon.com/blogs/industries/harnessing-the-power-of-generative-ai-for-automotive-technologies-on-aws/.
[5] *Artificial Intelligence in Government: Hearing Before the Senate Committee on Homeland Security and Government Affairs*, 118th Cong. (2023), https://www.hsgac.senate.gov/hearings/artificial-intelligence-in-government/.
[6] Karen Hao, *A New Vision of Artificial Intelligence for the People*, MIT Tech. Review (Apr. 22, 2022), https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/.

manipulative voice,[7] text,[8] and image synthesis.[9] Malicious actors have created "deepfakes"—including fake pornographic images and videos, particularly of women, without their consent.[10] Companies are deploying AI systems only to later recognize their dangers and recall them, often after the harm is already done. An eating disorder helpline chatbot was suspended after offering harmful advice to those struggling with recovery,[11] another for generating authoritative-sounding text on the benefits of eating glass.[12] A man took his own life after interacting with a chatbot, with his widow stating that "he would still be here" were it not for the AI.[13] A chatbot targeted at minors has been shown to generate harmful content,[14] and concerns about child safety led to a ban on one AI-powered "virtual friendship" service in Italy.[15] The output from generative AI can replicate damaging racist and sexist stereotypes.[16] Large language models can also "hallucinate," generating false content,[17] including potentially defamatory statements.[18]

Generative AI harms can also extend to the infrastructure used to create these models. The data centers on which AI depends can cause significant environmental harms. The amount of computer power put towards AI training — that is, teaching it to correctly interpret data and learn from it —is increasing exponentially,[19] with associated increases in energy consumption

---

[7] Pranshu Verma, *They Thought Loved Ones Were Calling For Help. It Was an AI Scam*, Wash. Post (Mar. 5, 2023), https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/.
[8] Jason Knowles & Ann Pistone, *Thieves Can Use ChatGPT to Write Convincing Scam Messages With Human-Like Language, Experts Warn*, ABC 7 Chicago (Mar. 14, 2023), https://abc7chicago.com/what-is-chatgpt-google-chatbot-ai-online-scams/12952645/.
[9] Hannah Gelbart, *Scammers Profit From Turkey-Syria Earthquake*, BBC News (Feb. 14, 2023), https://www.bbc.com/news/world-europe-64599553.amp.
[10] Coleman Spilde, *The College Student Whose Face Was Deepfaked Onto Porn*, Daily Beat (Mar. 11, 2023), https://www.thedailybeast.com/another-body-at-sxsw-the-college-girl-whose-face-was-deepfaked-onto-porn.
[11] Frances Vinall, *Eating-Disorder Group's AI Chatbot Gave Weight Loss Tips, Activist Says*, Wash. Post (June 1, 2023), https://www.washingtonpost.com/wellness/2023/06/01/eating-disorder-chatbot-ai-weight-loss/.
[12] Gerrick De Vynck et al., *Microsoft's AI Chatbot Is Going off the Rails*, Wash. Post (Feb. 16, 2023), https://www.washingtonpost.com/technology/2023/02/16/microsoft-bing-ai-chatbot-sydney/.
[13] Chloe Xiang, *'He Would Still Be Here': Man Dies by Suicide After Talking With AI Chatbot, Widow Says*, Motherboard (Mar. 30, 2023), https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says.
[14] Geoffrey A. Fowler, *Snapchat Tried to Make a Safe AI. It Chats With Me About Booze and Sex*, Wash. Post (Mar. 14, 2023), https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/.
[15] Natasha Lomas, *Replika, a 'Virtual Friendship' AI Chatbot, Hit With Data Ban in Italy Over Child Safety,* TechCrunch (Feb. 3, 2023), https://techcrunch.com/2023/02/03/replika-italy-data-processing-ban/.
[16] Melissa Heikkilä, *These New Tools Let You See for Yourself How Biased AI Image Models Are*, MIT Tech. Rev. (Mar. 22, 2023), https://www.technologyreview.com/2023/03/22/1070167/these-news-tool-let-you-see-for-yourself-how-biased-ai-image-models-are/; Davey Alba, *OpenAI Chatbot Spits Out Biased Musings, Despite Guardrails*, Bloomberg (Dec. 8, 2022), https://www.bloomberg.com/news/newsletters/2022-12-08/chatgpt-open-ai-s-chatbot-is-spitting-out-biased-sexist-results.
[17] Hussam Alkaissi & Samy I McFarlane, *Artificial Hallucinations in ChatGPT: Implications in Scientific Writing*, Nat'l. Library of Med. (Feb. 19, 2023), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9939079/.
[18] Lauren Leffer, *Australia Mayor Threatens to Sue OpenAI for Defamation by Chatbot*, Gizmodo (Apr. 5, 2023), https://gizmodo.com/openai-defamation-chatbot-brian-hood-chatgpt-1850302595.
[19] Jaime Sevilla et al., *Compute Trends Across Three Eras of Machine Learning,* ArXiv (Feb. 11, 2022), https://arxiv.org/abs/2202.05924.

and carbon dioxide emissions.[20] Accompanying these growing computing requirements are the life-cycle impacts of chip manufacture, from mining to e-waste.[21] Data center cooling can even stress local water supplies.[22]

Although advanced AI is often thought to depend purely on high-tech software engineers, it actually relies on a vast array of "ghost workers" across the world.[23] These workers are vital to the creation of large data sets used for training, providing feedback during model development to reduce harmful outputs, and monitoring the usage of models after their opening to public use. These workers are often employed under high pressure conditions, with low pay, and are exposed to disturbing and traumatizing outputs.[24]

Researchers have also highlighted potential risks from increasingly powerful AI, including risks that could result in widespread injury or death. For example, AI could help create chemical and biological weapons[25] or be used to launch nuclear weapons.[26] AI may be used to develop new hacking techniques, leading to vulnerabilities in core infrastructure.[27] AI-powered lethal autonomous weapons have already been developed,[28] and many researchers and industry figures have even raised the possibility of AI presenting an existential threat.[29] In a recent poll, more than one third of surveyed researchers said that AI development could lead to a "nuclear-level catastrophe."[30]

---

[20] Emma Strubell et al., *Energy and Policy Considerations for Deep Learning in NLP*, ArXiv (June 5, 2019), https://arxiv.org/abs/1906.02243.
[21] Anne-Laure Ligozat et al., *Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessments of AI Solutions*, MPDI (Apr. 25, 2022), https://www.mdpi.com/2071-1050/14/9/5172.
[22] David Mytton, *Data Centre Water Consumption*, npj Clean Water 4, 11 (Feb. 15, 2021), https://www.nature.com/articles/s41545-021-00101-w.
[23] Mary L Gray & Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*, HarperCollins (2019).
[24] Billy Perrigo, *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*, Time (Jan. 18, 2023), https://time.com/6247678/openai-chatgpt-kenya-workers/.
[25] Justine Calma, *AI Suggested 40,000 New Possible Chemical Weapons in Just Six Hours*, Verge (Mar. 17, 2023), https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx; Daniil A. Boiko et al., *Emergent Autonomous Scientific Research Capabilities of Large Language Models*, ArXiv (Apr. 11, 2023), https://arxiv.org/ftp/arxiv/papers/2304/2304.05332.pdf; Emily H. Soice et al., *Can large language models democratize access to dual-use biotechnology?*, ArXiv (June 6, 2023), https://arxiv.org/abs/2306.03809.
[26] Noah Greene, *AI Nuclear Weapons Catastrophe Can Be Avoided*, Lawfare (Mar. 2, 2023), https://www.lawfareblog.com/ai-nuclear-weapons-catastrophe-can-be-avoided.
[27] Daniel Oberhaus, *Prepare for AI Hackers*, Harvard Magazine (Mar. 2023), https://www.harvardmagazine.com/2023/03/right-now-ai-hacking.
[28] Robert Trager, *Killer Robots Are Here – And We Need to Regulate Them*, Foreign Policy (May 11, 2022), https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/.
[29] Derek Robertson, *Tracking the AI Apocalypse*, Politico (Jan. 10, 2023), https://www.politico.com/newsletters/digital-future-daily/2023/01/10/tracking-the-ai-apocalypse-00077279; Kevin Roose, *A.I. Poses 'Risk of Extinction,' Industry Leaders Warn*, New York Times (May 30, 2023), https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html.
[30] Tristan Bove, *A.I. Could Lead to a Nuclear-Level Catastrophe According to a Third of Researchers, a New Stanford Report Finds*, Fortune (Apr. 10, 2023), https://fortune.com/2023/04/10/ai-nuclear-level-catastrophe-experts-stanford/.

These current and potential future harms require urgent study. We ask GAO to assess this list of questions about harms from generative AI and potential strategies for mitigation. This is not a comprehensive list of possible harms, but addresses key areas identified by researchers and advocates in a rapidly evolving landscape.

1. To what extent do leading generative AI model providers generally follow key practices for generative AI training transparency, specifically for documenting and disclosing training data (including copyrighted and private consumer data), and model architectures?
2. What are the key practices for disclosures about testing and auditing procedures of generative AI algorithms? Do generative AI model developers follow these procedures?
3. What influence do commercial pressures, including the need to rapidly deploy products, have on the time allocated to pre-deployment testing of commercial models?
4. How can training data sets be protected against "data poisoning" (manipulating training data to cause harmful outputs of the trained model)?
5. Under what circumstances can training data, including potentially private or sensitive data, be extracted from generative AI models, and what options can be used to address this issue?
6. How can "prompt injection" or "jailbreaking" (specifically designing prompts for generative AI models that circumvent controls placed by the developer) be used to cause models to generate harmful output, and what options exist for combatting this?
7. What security measures do AI developers take to avoid their trained models being stolen by cyber attackers?
8. How do generative AI models rely on human workers for the process of data labelling and removing potentially harmful outputs? What are the impacts on these workers of being exposed to harmful material, and how can workers be protected from these harms?
9. What is known about potential harms of generative AI to various vulnerable populations, (for example, children, teens, those with mental health conditions, and those vulnerable to scams and fraud) and how are providers monitoring and mitigating such harms?
10. What are the current, and potential future environmental impacts, of large-scale generative AI deployment? This can include the impacts on energy consumption and grid stability, as well as the generation of e-waste, associated with the large-scale data-centers required to run AI models.
11. What is known about federal government agencies' research into and implementation of generative AI?

12. What frameworks are federal government agencies using to guide their use of generative AI?

13. What is known about the potential risks from increasingly powerful AI that could lead to injury, death, or other outcomes — up to human extinction — and how can such risks be addressed?

Given the rapid pace of development in this field, we would be happy to work with your staff to adjust the parameters of this study as you begin your work. Should you have any questions, please do not hesitate to contact our offices.

Sincerely,

Edward J. Markey
United States Senator

Gary C. Peters
United States Senator