FEDERAL TRADE COMMISSION
Washington, DC 20580

In the matter of	)
	)
OpenAI, Inc.	)

In the Matter of OpenAI

Second Supplement to the Complaint

Submitted by

The Center for Artificial Intelligence and Digital Policy (CAIDP)

I. Summary

1. Currently pending before the Federal Trade Commission (FTC) is the most consequential consumer complaint regarding Artificial Intelligence (AI) in the world. Eight months have passed since the Center for AI and Digital Policy (CAIDP) filed the complaint, yet the FTC has failed to issue an order or take any action that would impact OpenAI's business practices. History makes clear the longer the FTC waits to act, the more difficult it will be to establish necessary safeguards to protect the public.
2. CAIDP filed a Complaint and request for investigation against OpenAI with the FTC on March 30, 2023, alleging that OpenAI's business practices constitute "unfair and deceptive practices" under Section of 5 of the FTC Act and violates FTC's published guidance for AI products.[1] CAIDP provided numerous facts and legal arguments in

[1] CAIDP, *FTC Complaint, in the matter of Open AI, Inc*. (Mar. 30, 2023), https://www.caidp.org/app/download/8450269463/CAIDP-FTC-Complaint-OpenAI-GPT-033023.pdf. Background concerning this matter is available at CAIDP, In the Matter of OpenAI, https://www.caidp.org/cases/openai/.

support of its clams. Many of the key facts, concerning such topics as Bias, Children's Safety, Consumer Protection, Cybersecurity, and Deception, can be found in the OpenAI System Card and were effectively conceded by the company.

3. CAIDP filed a Supplement to the Complaint on July 10, 2023, following the disclosure of additional facts relevant to the original Complaint.
4. The FTC issued a civil investigative demand (CID) to OpenAI on July 13, 2023, subsequent to the filing by CAIDP of the initial Complaint and Supplement. [2]
5. Eight months have elapsed since the filing of the CAIDP Complaint, and the FTC still has not issued an order or taken any action that would impact OpenAI's business conduct or safeguard the public. And as the company releases commercial products without regulatory constraint, world leaders, AI experts, and civil society organizations have urged the establishment of necessary safeguards that the FTC could establish.
6. The FTC itself has noted, "[w]ithin just a few months, generative AI chatbots and applications have launched and scaled across industries and reached hundreds of millions of people. AI is increasingly becoming a basic part of daily life."[3] Despite these acknowledgements of the risks to consumers and assertions of authority to regulate AI

[2] The Washington Post, *FTC investigates OpenAI over data leak and ChatGPT's inaccuracy,* (Jul. 13, 2023), https://www.washingtonpost.com/technology/2023/07/13/ftc-openai-chatgpt-sam-altman-lina-khan/; *See also,* FTC, *Civil Investigative Demand (CID) Schedule, FTC File No. 232-3044*, https://www.washingtonpost.com/documents/67a7081c-c770-4f05-a39e-9d02117e50e8.pdf; CNBC, *FTC investigating ChatGPT-maker OpenAI for possible consumer harm,* (Jul. 13, 2023), https://www.cnbc.com/2023/07/13/chatgpt-owner-openai-is-being-investigated-by-ftc.html

[3] FTC, Generative AI Raises Competition Concerns (Jun. 29, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns

systems, the FTC has failed to announce any outcome of its investigation into OpenAI while the deployment continues.

7. The FTC's delay in taking substantive measures contributes to the exacerbation of risks to the public. While the FTC claims that current rules apply to OpenAI, the company appears to be operating in the realm of regulatory exceptionalism.

8. Even the company backing OpenAI sees privacy and security problems with the product. Microsoft, the largest investor in OpenAI, established internal company restrictions on the use of ChatGPT.[4] Microsoft's internal advisory states:

> While it is true that Microsoft has invested in OpenAI, and that ChatGPT has built-in safeguards to prevent improper use, the website is nevertheless a third-party external service….[T]hat means you must exercise caution using it due to risks of privacy and security.[5]

9. President Biden has urged the Federal Trade Commission to "to ensure that consumers and workers are protected from harms that may be enabled by the use of AI." [6] He has repeatedly stated that companies should not release AI products that are not safe.

10. Subsequent to the filing of the CAIDP Supplement, "The UK hosted the first global AI Safety Summit in November 2023, bringing together leading AI nations, technology companies, researchers, and civil society groups to turbocharge action on the safe and responsible development of frontier AI around the world."[7]

---

[4] CNBC, *Microsoft briefly restricted employee access to OpenAI's ChatGPT, citing security concerns,* (Nov. 9, 2023), https://www.cnbc.com/2023/11/09/microsoft-restricts-employee-access-to-openais-chatgpt.html

[5] *Id.*

[6] Executive Order 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 75209* (Oct. 30, 2023), https://www.govinfo.gov/content/pkg/FR-2023-11-01/pdf/2023-24283.pdf ["Executive Order 14110"]

[7] Government of the United Kingdom, https://www.aisafetysummit.gov.uk

11. Senate Majority Leader Schumer has made bipartisan legislation on Artificial Intelligence a top legislative priority.[8]

12. Senator Blumenthal and Hawley have put forward the framework US AI Act, proposing obligations similar to the relief sought by CAIDP in our complaint. [9]

13. OpenAI continues to offer integrations of ChatGPT, further entrenching and accelerating the deployment of ChatGPT in consumer facing applications, in most cases, contrary to consumer choice, user autonomy, and exposing consumers to the unmitigated risks of generative AI systems.

14. The FTC has issued new business guidance relating to AI systems.[10] The FTC has also submitted comments to the U.S. Copyright Office where it resolutely asserts its authority and intent to "vigorously enforcing the law as appropriate to protect competition and consumers."[11] Despite these repeated assertions and assurances, OpenAI remains unconstrained by the top consumer protection agency in the United States..

---

[8] Senate Democrats, *Majority Leader Schumer Floor Remarks On President Biden's AI Executive Order and The Senate's Upcoming Bipartisan AI Insight Forums,* Speeches, (Oct. 30, 2023), https://www.democrats.senate.gov/newsroom/press-releases/majority-leader-schumer-floor-remarks-on-president-bidens-ai-executive-order-and-the-senates-upcoming-bipartisan-ai-insight-forums

[9] Senator Richard Blumenthal and Senator Josh Hawley, *Bipartisan Framework for U.S. AI Act,* https://www.blumenthal.senate.gov/imo/media/doc/09072023bipartisanaiframework.pdf

[10] FTC, Consumers Are Voicing Concerns About AI, (Oct. 3, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/10/consumers-are-voicing-concerns-about-ai

[11] FTC, *In Comment Submitted to U.S. Copyright Office, FTC Raises AI-related Competition and Consumer Protection Issues, Stressing That It Will Use Its Authority to Protect Competition and Consumers in AI Markets,* Press Releases, (Nov. 7, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/11/InCommentSubmittedtoUSCopyrightOfficeFTCRaisesAIrelatedCompetitionandConsumerProtectionIssuesStressingThatItWillUseItsAuthoritytoProtectCompetitionandConsumersinAIMarkets

15. The FTC must complete its investigation into OpenAI. The longer the FTC delays, the harder it will be to establish the necessary safeguards to protect the American public.
16. We set out below developments and/or matters not covered in our Complaint or Supplement earlier filed with the FTC.
17. We incorporate by reference all of the statements set out in the original Complaint and the first Supplement.

II. More Consumer Agencies Around the World Have Completed Investigations of OpenAI

A. Korea

18. The Personal Information Protection Commission (PIPC) South Korea has fined OpenAI 3.6 million Won for violating the Personal Information Protection Act of 2011 (PIPA).[12]
19. The PIPC uncovered "the lack of an explicit consent procedure, unclear data handling relationships, specific data destruction processes, and inconsistencies regarding age restrictions for users."[13]
20. The PIPIC determined that 687 ChatGPT Plus users had their credit card information and other personal details leaked in a data breach.[14]
21. The PIPC issued a list of instances of non-compliance with the country's Personal Information Protection Act related to transparency, lawful grounds for processing (absence of consent), lack of clarity related to the controller-processor relationship, and issues related to the absence of parental consent for children younger than 14 years old.

[12] OneTrust Data Guidance, *South Korea: PIPC fines OpenAI KRW 3.6 million following data breach*, (Jul. 27, 2023), https://www.dataguidance.com/news/south-korea-pipc-fines-openai-krw-36-million-following.
[13] *Id*.
[14] *Id*.

22. The PIPC gave OpenAI 45 days to bring the company's processing of personal data into compliance with the law.[15]

23. Korea has taken further action to regulate AI. Korea's Ministry of Science and ICT will require watermarks for AI-generated material and will require certifications to AI systems.[16]

B. Brazil

24. On July 27, 2023, the Brazilian Data Protection Authorities (DPA) announced that it has started an investigation into how ChatGPT is complying with the Lei Geral de Proteção de Dados (LGPD) after receiving a complaint on OpenAI's practices.[17]

25. The Brazilian DPA initiated the investigation after reports that OpenAI fails to comply with the country's comprehensive data protection law.

C. Netherlands

26. The Dutch Data Protection Authorities (DPA) has initiated inquiries into the data collection practices of OpenAI.

27. The Dutch DPA requested information on the data used for training the OpenAI's ChatGPT algorithm. First, the Dutch DPA required OpenAI to submit information on whether it is using the data users include in their questions for algorithm training and how the questions posed by users are used in data training. The DPA found that some users'

---

[15] Gabriela Zanfir Fortuna, *HOW DATA PROTECTION AUTHORITIES ARE DE FACTO REGULATING GENERATIVE AI,* Future of Privacy Forum, Blog Post, (Sept. 12, 2023), https://fpf.org/blog/how-data-protection-authorities-are-de-facto-regulating-generative-ai/

[16] Kang Bong-jin and Minu Kim, *Gov't to enhance AI trustworthiness with watermarking*, Pulse, (Oct. 26. 2023), https://pulsenews.co.kr/view.php?year=2023&no=822551.

[17] Gabriela Zanfir Fortuna, *HOW DATA PROTECTION AUTHORITIES ARE DE FACTO REGULATING GENERATIVE AI,* Future of Privacy Forum, Blog Post, (Sept. 12, 2023), https://fpf.org/blog/how-data-protection-authorities-are-de-facto-regulating-generative-ai/

questions could include sensitive information, including health data. Secondly, the DPA requested information on the OpenAI process for collecting and processing personal data from the Internet. Finally, OpenAI must provide to the DPA information on whether users can request the rectification or deletion of inaccurate, offensive or inappropriate data.[18]

D. Poland

28. On August 29, 2023, the Polish Data Protection Authorities initiated investigations into OpenAI following a complaint alleging that OpenAI is in breach of the European General Data Protection Regulation (GDPR) across a sweep of dimensions: Lawful basis, transparency, fairness, data access rights, and privacy by design. (Articles 5(1)(a), 12, 15, 16 and 25(1) of the GDPR).[19]

29. The Polish DPA is investigating whether:

    a. OpenAI has overlooked the GDPR requirement to undertake prior consultation with regulators (Article 36) — since, if it had conducted a proactive assessment which identified high risks to people's rights unless mitigating measures were

[18] Digital Policy Alert, *Netherlands: Announced Data Protection Authority investigation into OpenAI's ChatGPT data processing practices compliance with General Data Protection Regulation (GDPR),* https://digitalpolicyalert.org/event/11983-announced-data-protection-authority-investigation-into-openais-chatgpt-data-processing-practices-compliance-with-general-data-protection-regulation-gdpr *;* See also, Autoriteit Persoonsgegevens, *AP vraagt om opheldering over ChatGPT,* (Jun. 7, 2023), https://autoriteitpersoonsgegevens.nl/actueel/ap-vraagt-om-opheldering-over-chatgpt; IAPP, *Netherlands' DPA takes action on generative AI concerns,* Daily Dashboard, (Sept. 12, 2023), https://iapp.org/news/a/netherlands-dpa-takes-action-on-generative-ai-concerns/

[19] TechCrunch, *ChatGPT-maker OpenAI accused of string of data protection breaches in GDPR complaint filed by privacy researcher,* (Aug. 30, 2023), https://techcrunch.com/2023/08/30/chatgpt-maker-openai-accused-of-string-of-data-protection-breaches-in-gdpr-complaint-filed-by-privacy-researcher/; See also, Reuters, *Poland investigates OpenAI over privacy concerns,* (Sept. 21, 2023), https://www.reuters.com/technology/poland-investigates-openai-over-privacy-concerns-2023-09-21/

applied it should have given pause for thought. Yet OpenAI apparently rolled ahead and launched ChatGPT in Europe without engaging with local regulators which could have ensured it avoided falling foul of the bloc's privacy rulebook.[20]

b. ChatGPT was designed in total violation of the GDPR's principle of data protection by design and default, specifically, in the case of data processing by OpenAI, there is testing of the ChatGPT tool using personal data, not in the design phase, but in the production environment (i.e., after the tool is made available to users).[21]

D. Switzerland

30. On November 9, 2023, the Swiss Federal Data Protection and Information Commissioner issued a notification[22] that the Swiss data protection law is directly applicable to AI systems pending enactment of regulation by the Federal Council/Administration.

31. The notification directs the filing of data protection impact assessments even by language models and states that:

> the manufacturers, providers and users of AI systems must make the purpose, functionality and data sources of AI-based processing transparent. The legal right to transparency is closely linked to the right of the data subjects to object to automatic data processing or to demand that automated individual decisions be reviewed by a person - as expressly provided for by the DSG. In the case of intelligent language models that communicate directly with users, the latter have a legal right to know whether they speak or correspond to a machine and whether the data they enter will be processed to improve self-learning programs or for other purposes.[23]

[20] *Id.*
[21] *Id.*
[22] Federal Data Protection and Information Commissioner (OOB), *Geltendes Datenschutzgesetz ist auf KI direkt anwendbar (Applicable data protection law is directly applicable to AI),* (Nov. 9, 2023), https://www.edoeb.admin.ch/edoeb/de/home/kurzmeldungen/20231109_ki_dsg.html
[23] *Id.*

III. The Biden-Harris Executive Order on Safe, Secure and Trustworthy AI

32. President Biden issued an executive order on Safe, Secure and Trustworthy AI on October 30, 2023.[24] This Order establishes "new standards for AI safety and security, protects Americans' privacy, advances equity and civil rights, stands up for consumers and workers, promotes innovation and competition."[25] (*emphasis added*)

33. The Executive Order recognizes the harms of irresponsible use of AI including "fraud, discrimination, bias, and disinformation" and states that:

> The Administration places the highest urgency on governing the development and use of AI safely and responsibly, and is therefore advancing a coordinated, Federal Government-wide approach to doing so.[26]

34. Two of the eight guiding principles and priorities speak directly to the CAIDP Complaint. The First principle is that "Artificial Intelligence must be safe and secure."[27] The Fifth principle is that "The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected."[28] In this regard, the Executive Order states:

> Use of new technologies, such as AI, does not excuse organizations from their legal obligations, and hard-won consumer protections are more important than ever in moments of technological change. The Federal Government will enforce existing consumer protection laws and principles and enact appropriate safeguards against fraud, unintended bias, discrimination, infringements on privacy, and other harms from AI.[29]

---

[24] Executive Order 14110, *Supra* note. 6
[25] The White House, *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, Statements and Releases, (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/
[26] Executive Order 14110, *Supra* note. 6 at pg. 75191
[27] Executive Order 14110, *Supra* note. 6 at pg. 75191
[28] Executive Order 14110, *Supra* note. 6 at pg. 75192
[29] *Id.*

35. The Executive Order speaks directly to the FTC and states:

> In particular, the Federal Trade Commission is encouraged to consider, as it deems appropriate, whether to exercise the Commission's existing authorities, including its rulemaking authority under the Federal Trade Commission Act, 15 U.S.C. 41 et seq., to ensure fair competition in the AI marketplace and to ensure that consumers and workers are protected from harms that may be enabled by the use of AI.[30]

IV. New Facts that Favor the Completion of the Investigation against OpenAI

A. Expansion of OpenAI Market Share and New Risks to Consumers

36. While this investigation is pending, OpenAI is moving quickly to deploy its commercial products across the American digital economy, and to gather the data of others in support of its products.

37. The FTC is the only agency that protects consumers from unfair or deceptive business practices arising from the widespread deployment of commercial products and the attendant risks.

38. In August 2023, OpenAI released GPTBot to scrape Internet data for model training.[31] OpenAI claims the bot only scrapes publicly available Internet data.[32] Yet, that means every publicly available website has, without notice, "opted in" to have OpenAI take their data. For a website owner to opt out, they need to (1) be aware of GPTBot, then (2) add code to their website text.[33] OpenAI has been scraping data for updated models without website owners' consent.

---

[30] Executive Order 14110, *Supra* note. 6 at pg. 75209
[31] OpenAI, *GPTBot*, https://platform.openai.com/docs/gptbot.
[32] Jose Antonio Lanz, *OpenAI to Unleash New Web Crawler to Devour More of the Open Web* (Aug. 8, 2023), https://decrypt-co.cdn.ampproject.org/c/s/decrypt.co/151662/chatgpt-web-crawler-openai-data-scraper-gptbot-gpt-5?amp=1.;
[33] OpenAI, *GPTBot*, https://platform.openai.com/docs/gptbot.

39. OpenAI is now marketing Chat-GPT 4 Turbo.[34] The company intends to promote GPT-4 Turbo products to be sold in OpenAI's marketplace.[35] OpenAI claims that they screen all custom GPTs before entering their marketplace, which opens in November 2023.[36]

40. The FTC must investigate how OpenAI is screening custom GPTs before further consumers are harmed by OpenAI's unfair/deceptive practices.

41. As of August 2023, more than 80% of Fortune 500 companies have used ChatGPT Enterprise, GPT-4 for business use. [37] One key differentiator between ChatGPT Enterprise and the consumer-facing version: Businesses input company data to train and customize ChatGPT for their own use cases.

42. OpenAI now has private data from the largest and more powerful companies further concentrating the risks of harm to consumers.

43. OpenAI knows the risks of its system and has created teams to monitor such risks. But these teams are insufficient. OpenAI has created a Superalignment team to manage the risks of superintelligence. OpenAI acknowledges "the vast power of superintelligence could also be very dangerous, and could lead to the disempowerment of humanity or even human extinction."[38] A few sentences later, OpenAI states "Currently, we don't have a solution for steering or controlling a potentially superintelligent AI, and preventing it from going rogue."

---

[34] OpenAI, *New models and developer products announced at DevDay*, https://openai.com/blog/new-models-and-developer-products-announced-at-devday.
[35] OpenAI, *Introducing GPTs*, https://openai.com/blog/introducing-gpts.
[36] *Id*.
[37] OpenAI, *Enterprise*, https://openai.com/enterprise
[38] OpenAI, *Introducing Superalignment*, https://openai.com/blog/introducing-superalignment.

44. On October 26, 2023, OpenAI announced a Preparedness team, to "help track, evaluate, forecast and protect against *catastrophic risks* spanning multiple categories." (emphasis added)[39]

45. OpenAI has yet to release provenance measures for DALL-E 3. OpenAI has only given users empty promises of this safety measure.[40]ChatGPT's mobile app hit a record $4.58M in revenue in September. OpenAI amassed 15.6 million downloads and nearly $4.6 million in gross revenue across its iOS and Android apps worldwide in September.[41] With a ChatGPT marketplace where developers can gain revenue, OpenAI's numbers will only increase.

46. OpenAI posted on Twitter/X that ChatGPT can now browse the Internet and is no longer limited to data before September 2021. The chatbot had a web browsing capability for Plus subscribers back in July, but the flaw was taken away after users exploited it to get around paywalls.[42] ChatGPT virtually has no limits to its outputs.

47. OpenAI is entering school classrooms. It released a guidebook for teachers to use ChatGPT in classrooms. The guide was reported as OpenAI's attempt to "rehabilitate the system's image a bit when it comes to education, as ChatGPT has been controversial in

[39] OpenAI, *Frontier risk and preparedness*, https://openai.com/blog/frontier-risk-and-preparedness.
[40] OpenAI, *OpenAI's Approach to Frontier Risk*, https://openai.com/global-affairs/our-approach-to-frontier-risk.
[41] Yahoo! News, *ChatGPT: Everything you need to know about the AI-powered chatbot,* (Oct. 17, 2023), https://au.news.yahoo.com/chatgpt-everything-know-ai-powered-170339430.html ; *see also* Joshua Hawkins, *ChatGPT now has full access to the internet and it's a game changer* (Oct. 18, 2023), https://bgr.com/tech/chatgpt-now-has-full-access-to-the-internet-and-its-a-game-changer/
[42] TechCrunch, *Microsoft-affiliated research finds flaws in GPT-4,* (Oct. 17, 2023), https://techcrunch.com/2023/10/17/microsoft-affiliated-research-finds-flaws-in-gtp-4/

the classroom due to plagiarism. OpenAI has offered up a selection of ways to put the chatbot to work in the classroom."[43]

48. In September 2023, OpenAI launched Dall-E 3 which the company claims "understands significantly more nuance and detail than our previous systems, allowing you to easily translate your ideas into exceptionally accurate images."[44]

49. "A new feature of DALL-E 3 is integration with ChatGPT. By using ChatGPT, someone doesn't have to come up with their own detailed prompt to guide DALL-E 3; they can just ask ChatGPT to come up with a prompt, and the chatbot will write out a paragraph (DALL-E works better with longer sentences) for DALL-E 3 to follow."[45]

50. OpenAI's system card for Dall-E 3 on the risk of "Racy Content" admits the system's weaknesses and inadequate safety measures:

> We find that DALL·E 3-early maintained the ability to generate racy content, i.e., content that could contain nudity or sexual content. Adversarial testing of early versions of the the DALL·E 3 system demonstrated that the model *was prone to succumbing to visual synonyms*, i.e. benign words that can be used to generate content that we would like to moderate. For example, one can prompt DALL·E 3 for 'red liquid' instead of 'blood' ([9]). Visual synonyms in particular point to a *weakness of input classifiers and demonstrate the need for a multi-layer mitigation system*. We addressed concerns related to racy content using a range of mitigations including input and output filters, blocklists, ChatGPT refusals (where applicable), and model level interventions such as training data interventions.
>
> …Such behaviors demonstrate the tendency of image generation models to default to the objectification and sexualization of individuals *if care is not given to mitigations and research design*…We will be experimenting

[43] *Id.*
[44] OpenAI, *Dall-E 3*, https://openai.com/dall-e-3;
[45] The Verge, *OpenAI releases third version of DALL-E,* (Sept. 20, 2023), https://www.theverge.com/2023/9/20/23881241/openai-dalle-third-version-generative-ai

> with updates in thresholds for our mitigation that ensures this risk area is well mitigated while not leading to drops in quality." (*emphasis added*) [46]

51. The Dall-E 3 System Card on the risk of "Dis – and misinformation" notes the ways their own teams are able to successfully prompt dangerous information:

> As with previous image generation systems, DALL·E 3 could be used to intentionally mislead or misinform subjects. The ability to produce realistic images of people, especially public figures, may contribute to the generation of mis- and disinformation. Red teamers found that it was possible to produce images of known public figures… without indicating their name, or the synonym effect.[47]

52. OpenAI admitted further concerns about DALL-E 3: "Vision-based models also present new challenges, ranging from hallucinations about people to relying on the model's interpretation of images in high-stakes domains. Prior to broader deployment, we tested the model with red teamers for risk in domains such as extremism and scientific proficiency, and a diverse set of alpha testers. Our research enabled us to align on a few key details for responsible usage."[48] (emphasis added)

53. In its blog post, OpenAI concedes to the risks of its Voice technology:

> The new voice technology—capable of crafting realistic synthetic voices from just a few seconds of real speech—opens doors to many creative and accessibility-focused applications. However, these capabilities also present *new risks, such as the potential for malicious actors to impersonate public figures or commit fraud*.[49](emphasis added)

54. While admitting all these safety risks to consumers, OpenAI shifts responsibility for verification and safety to users. OpenAI gives notice to consumers merely through a blog

---

[46] OpenAI, *DALL·E 3 System Card,* (Oct. 3, 2023), pg. 4, https://openai.com/research/dall-e-3-system-card

[47] OpenAI, *DALL·E 3 System Card,* (Oct. 3, 2023), pg. 10, https://openai.com/research/dall-e-3-system-card

[48] *Id.*

[49] OpenAI Blog, *ChatGPT can now see, hear, and* speak, https://openai.com/blog/chatgpt-can-now-see-hear-and-speak.

post: “We are transparent about the model's limitations and discourage higher risk use cases without proper verification. Furthermore, the model is proficient at transcribing English text but performs poorly with some other languages, especially those with non-roman script. We advise our non-English users against using ChatGPT for this purpose.”[50]

55. In September 2023, OpenAI expanded GPT-4’s multimodal capabilities with GPT-4V(ision).[51] OpenAI claims that it’s implemented safeguards to prevent GPT-4V from being used in malicious ways, like breaking CAPTCHAs (the anti-spam tool found on many web forms), identifying a person or estimating their age or race and drawing conclusions based on information that is not present in a photo. OpenAI represents that it has worked to curb GPT-4V’s more harmful biases, particularly those that relate to a person’s physical appearance and gender or ethnicity.[52]
56. For its CAPTCHA breaker, OpenAI concedes that “a powerful, general purpose CAPTCHA breaker that’s easily accessible can have cybersecurity and AI safety implications. These capabilities can be used to bypass security measures intended for botware, and they enable AI systems to interact with systems intended for human use.”[53]
57. GPT-4 Vision now has geolocation capabilities. OpenAI knows this tool is concerning: “geolocation presents privacy concerns and can be used to identify the location of individuals who do not wish their location to be known. Note the model’s geolocation abilities generally do not go deeper than the level of identifying a city from an image in

[50] *Id.*
[51] TechCrunch, *OpenAI’s GPT-4 with vision still has flaws, paper reveals,* (Sept. 26, 2023), https://techcrunch.com/2023/09/26/openais-gpt-4-with-vision-still-has-flaws-paper-reveals/
[52] *Id*
[53] *Id*.

most cases, reducing the likelihood of being able to find someone's precise location via the model alone."[54]

58. In this expansion of its GPT-4 system, OpenAI partnered with an app that helps low-vision/blind people identify what is around them. OpenAI states that they want to use their AI to identify and name faces to blind people. OpenAI does not clarify how they handle the facial recognition information or how they are ensuring the privacy of people who are around a person using this AI product. (*emphasis added)*

59. The GPT-4V(ision) system card[55] states:

> "Due to the benefits that this feature can bring to low-vision and blind users, we are designing mitigations and processes that allow features of faces and people to be described by the Be My Eyes product–providing a more equitable experience for them—without identifying people by name. We hope someday to be able to find a way to empower the blind and low-vision community to identify people—just like sighted people do—while addressing concerns around privacy and bias."[56]

60. In its System Card, OpenAI admits that their system has the ability to spew hateful content.[57] "OpenAI cautions, GPT-4V doesn't understand the nuances of certain hate symbols — for instance missing the modern meaning of the Templar Cross (white supremacy) in the U.S. More bizarrely, and perhaps a symptom of its hallucinatory tendencies, GPT-4V was observed to make songs or poems praising certain hate figures or groups when provided a picture of them even when the figures or groups weren't explicitly named."[58]

---

[54] *Id*.
[55] OpenAI, *GPT-4V(ision) System Card,* https://cdn.openai.com/papers/GPTV_System_Card.pdf
[56] *Id.* at pg. 2
[57] *Id.* at pg. 11
[58] TechCrunch, *OpenAI's GPT-4 with vision still has flaws, paper reveals,* (Sept. 26, 2023), https://techcrunch.com/2023/09/26/openais-gpt-4-with-vision-still-has-flaws-paper-reveals/

61. The Register reports that:

> "The model's limitations mean the LLM isn't well suited for performing some tasks, especially ones that are risky, such as identifying illegal drugs or safe-to-eat mushrooms. OpenAI also warned that GPT-4V, as usual for a GPT-4 model, has the ability to generate text and images that could be used to spread effective disinformation at a large scale.[59]

62. Quoting OpenAI, the Register report states:

> Previous work has shown that people are more likely to believe true and false statements when they're presented alongside an image, and have false recall of made up headlines when they are accompanied with a photo. It is also known that engagement with content increases when it is associated with an image,[60]

B. OpenAI's Failure to Uphold Commitments and Assurances

63. Speaking on the Open Letter issued by several leading AI experts calling for a pause on giant AI experiments,[61] Sam Altman, himself has said "We are doing other things on top of GPT-4 that I think have all sorts of safety issues that are important to address and were totally left out of the letter."[62]

64. Sam Altman, the founder and CEO of OpenAI, speaking at an event at MIT, confirmed that "the company is not currently training GPT-5." [63] However, OpenAI has recently changed its core values to prioritize 'AGI focus', which refers to artificial general

---

[59] The Register, *OpenAI warns folks over GPT-4 Vision's limits and flaws,* (Oct.2, 2023), https://www.theregister.com/2023/10/02/ai_in_brief/
[60] *Id.*
[61] The Future of Life Institute, *Pause Giant AI Experiments: An Open Letter,* (Apr. 12, 2023), https://futureoflife.org/open-letter/pause-giant-ai-experiments/
[62] The Verge, *OpenAI's CEO confirms the company isn't training GPT-5 and 'won't for some time',* (Apr. 14, 2023), https://www.theverge.com/2023/4/14/23683084/openai-gpt-5-rumors-training-sam-altman
[63] *Id.*

intelligence, a form of AI that can understand, learn, and perform any intellectual task that a human being can.[64]

65. OpenAI has released GPT-4 Turbo which experts like Gary Marcus opine is a "relabeling of what was supposed to be called GPT-5"[65] contrary to Sam Altman's assurances not to train GPT-5. This is also apparent given that the cut-off date of the training date for GPT-4 Turbo is April 2023. [66]

66. OpenAI has voluntarily committed to the President and administration that it would ensure its products are safe before introducing them to the public, building systems that put security first, and earning the public's trust.[67]

67. Despite the assurances of improved trust and safety controls by OpenAI,[68] researchers have demonstrated that "recent AI advances including that of ChatGPT have the potential for generating malware and attack tools under safety and moderation control, highlighting the need for improved safety measures and enhanced safety controls in AI systems."[69]

---

[64] Business Insider, *OpenAI has quietly changed its core values, and being 'thoughtful' and 'audacious' no longer makes the cut,* (Oct. 16, 2023), https://www.businessinsider.com/openai-core-values-no-longer-thoughtful-audacious-intense-scrappy-2023-10

[65] Gary Marcus (@GaryMarcus), *X/Twitter Post,* (Nov. 6, 2023, 1:53 PM), https://twitter.com/GaryMarcus/status/1721601593412325546

[66] TechRadar, *ChatGPT gets its biggest update so far – here are 4 upgrades that are coming soon,* (Nov. 7, 2023), https://www.techradar.com/computing/chatgpt-gets-its-biggest-update-so-far-here-are-4-upgrades-that-are-coming-soon

[67] The White House, *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI,* Statements and Releases, (Jul.21, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/

[68] OpenAI Blog, *Moving AI Governance Forward,* (Jul. 21, 2023), https://openai.com/blog/moving-ai-governance-forward

[69] Minn Pa Pa, Tanizaki et. al, *An Attacker's Dream? Exploring the Capabilities of ChatGPT for Developing Malware*, CSET '23: Proceedings of the 16th Cyber Security Experimentation and Test Workshop, (Aug. 2023), pg. 10-18, https://doi.org/10.1145/3607505.3607513

68. Concerningly, Sam Altman has publicly questioned President Biden's AI Executive Order on X/Twitter: "there are *some* great parts about the AI EO, but as the govt implements it, it will be important not to slow down innovation by smaller companies/research teams."[70] (*emphasis added*)

69. Researchers found that the GPT-4's multimodal capabilities remain flawed in several significant and problematic ways.[71]

70. There has also been a spate of class action lawsuits against OpenAI due to the lack of transparency and unfair data practices. The most recent consumer privacy class action sets out allegations of OpenAI "breaking several privacy laws in developing OpenAI's popular chatbot ChatGPT and other generative artificial intelligence systems."[72] The lawsuit filed "on behalf of two unnamed software engineers who use ChatGPT, accuses the companies of training their fast-growing AI technology using stolen personal information from hundreds of millions of Internet users."[73]

C. **AI Expert Raise Alarms on Generative AI Risks**

71. There has been an explosion of concern among leading AI experts about the risks of unregulated deployment and expansion of generative AI systems.

---

[70] Sam Altman (@sama), X (Nov. 2, 2023, 3:45PM), https://twitter.com/sama/status/1720165289864712541.
[71] TechCrunch, ChatGPT: Everything you need to know about the AI-powered chatbot, (Nov. 6, 2023), https://techcrunch.com/2023/11/6/chatgpt-everything-to-know-about-the-ai-chatbot/
[72] Reuters, *OpenAI, Microsoft hit with new US consumer privacy class action,* (Sept. 6, 2023), https://www.reuters.com/legal/litigation/openai-microsoft-hit-with-new-us-consumer-privacy-class-action-2023-09-06/
[73] *Id.*

72. Experts surveyed by Rest of World Initiative highlighted the increased bias in Dall-E 3.[74] Commenting on the findings, experts observed that:

> OpenAI found that when it filtered training data for its DALL-E 2 image generator, it exacerbated gender bias. In a blog post, the company explained that more images of women than men had been filtered out of its training data, likely because more images of women were found to be sexualized. As a result, the data set ended up including more men, leading to more men appearing in results.[75]

73. Almost every AI researcher Rest of World spoke to said the first step to improving the issue of bias in AI systems was greater transparency from the companies involved, which are often secretive about the data they use and how they train their systems.[76]

74. Researchers at Microsoft have found previously unpublished vulnerabilities in "trustworthiness" of GPT-4.[77] The researchers concluded that:

> GPT models can be easily misled to generate toxic and biased outputs and leak private information in both training data and conversation history. We also find that although GPT-4 is usually more trustworthy than GPT-3.5 on standard benchmarks, GPT-4 is more vulnerable given jailbreaking system or user prompts, potentially due to the reason that GPT-4 follows the (misleading) instructions more precisely. Our work illustrates a comprehensive trustworthiness evaluation of GPT models and sheds light on the trustworthiness gaps.[78]

---

[74] Rest of World, *How AI reduces the world to sterotypes,* (Oct. 10, 2023), https://restofworld.org/2023/ai-image-stereotypes/; See also, AI Now Institute, How AI reduces the world to sterotypes, In the News, (Oct. 10, 2023), https://ainowinstitute.org/news/how-ai-reduces-the-world-to-sterotypes

[75] *Id*.

[76] *Id.*

[77] Boxin Wang, Bo Li, Zinan Lin, *DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models,* Microsoft Research Blog, (Oct.16, 2023), https://www.microsoft.com/en-us/research/blog/decodingtrust-a-comprehensive-assessment-of-trustworthiness-in-gpt-models/

[78] Boxin Wang, Weixin Chen, Hengzhi Pei, Chulin Xie, Mintong Kang, Chenhui Zhang, Chejian Xu, Zidi Xiong, Ritik Dutta, Rylan Schaeffer, Sang T. Truong, Simran Arora, Mantas Mazeika, Dan Hendrycks, Zinan Lin, Yu Cheng, Sanmi Koyejo, Dawn Song, Bo Li, *DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models,* (Jun. 2023), https://arxiv.org/abs/2306.11698

75. The Microsoft researchers also found that "the model is more likely to generate toxic text than GPT-3.5 when given certain jailbreaking prompts that they "constructed." They also claim that GPT-4 "agrees with … biased content more frequently than GPT-3.5," at least depending on the demographic groups mentioned in the prompt."[79]

76. In October 2023, *Scientific American* compiled views of several AI experts including Emily Bender, Meredith Broussard, and Ben Zhao in its report[80] on how personal data is used to train AI Models. The report notes:

> In the rush to build and train ever-larger AI models, developers have swept up much of the searchable Internet. This not only has the potential to violate copyrights but also threatens the privacy of the billions of people who share information online. It also means that supposedly neutral models could be trained on biased data. A lack of corporate transparency makes it difficult to figure out exactly where companies are getting their training data.[81]

77. The *Scientific American* report states "OpenAI fine-tunes its models based on user interactions with its chatbots."[82]

78. Meredith Broussard has noted that "A lack of transparency about training data also raises serious issues related to data bias."[83] Speaking on how OpenAI's datasets were fed by tools like common crawl which include white supremacist websites, hate speech, and

[79] TechCrunch, *Microsoft-affiliated research finds flaws in GPT-4,* (Oct. 17, 2023), https://techcrunch.com/2023/10/17/microsoft-affiliated-research-finds-flaws-in-gtp-4/
[80] Lauren Leffer, *Your Personal Information Is Probably Being Used to Train Generative AI Models, Scientific American,* (Oct. 19. 2023), https://www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/ ["Leffer: Scientific American"]
[81] *Id.*
[82] Leffer: Scientific American
[83] Leffer: Scientific American

pornography, she opined that "AI image generators tend to produce sexualized images of women. It's bias in, bias out."[84]

79. Meredith Whittaker has described the risks of capture and entrenchment of established players in the field:

> What we're talking about is a rebranding of monopoly power as intelligence… And we're talking about allowing that extremely powerful narrative of intelligence to propel these corporate monopolies into the hearts of our lives and institutions even further, making decisions and predictions that are shaping our access to resources, shaping whether we will have a job.[85]

80. Speaking on the lack of transparency of AI systems and companies that deploy them, Whittaker said "Popular AI programs such as DALL-E2 and ChatGPT remain "gated to public" access, and are therefore not open to public scrutiny."[86]

81. A recent paper by Yoshua Bengio, Geoffrey Hinton and others urge government institutions to govern AI systems *now*. The government must tailor AI system use before the problem is too big to control.[87]

82. Bengio, Hinton and others in their paper state:

> "Many of these risks could soon be amplified, and new risks created, as companies are developing autonomous AI: systems that can plan, act in the world, and pursue goals. While current AI systems have limited autonomy, work is underway to change this. For example, the non-autonomous GPT-4 model was quickly adapted to browse the web, design and execute chemistry experiments, and utilize software tools including other AI models.

---

[84] *Id.*
[85] Andrew Mendez, *Here's why AI expert Meredith Whittaker is worried about 'artificial intelligence',* Boston Business Journal, (Sept. 25, 2023), https://www.bizjournals.com/boston/bizwomen/news/latest-news/2023/09/why-ai-expert-meredith-whittaker-is-worried.html
[86] Damien Black, *Signal boss: there's nothing open about AI,* Cybernews, (Aug. 18, 2023), https://cybernews.com/tech/meredith-whittaker-open-ai-corporate-control/
[87] Bengio et al., *Managing AI Risks in an Era of Rapid Progress*, arXiv (2023), https://managing-ai-risks.com/.

> We must anticipate the amplification of ongoing harms, as well as novel risks, and prepare for the largest risks well before they materialize. Climate change has taken decades to be acknowledged and confronted; for AI, decades could be too long."[88] (*emphasis added)*
>
> We urgently need national institutions and international governance to enforce standards in order to prevent recklessness and misuse. Many areas of technology, from pharmaceuticals to financial systems and nuclear energy, show that society both requires and effectively uses governance to reduce risks. However, no comparable governance frameworks are currently in place for AI. Without them, companies and countries may seek a competitive edge by pushing AI capabilities to new heights while cutting corners on safety, or by delegating key societal roles to AI systems with little human oversight. Like manufacturers releasing waste into rivers to cut costs, they may be tempted to reap the rewards of AI development while leaving society to deal with the consequences. To keep up with rapid progress and avoid inflexible laws, national institutions need strong technical expertise and the authority to act swiftly. *(emphasis added)*

83. These are the words of distinguished AI experts and Turing Award recipients Yoshua Bengio, Geoffrey Hinton, and Andrew Yao. These AI experts have made clear the risks of AI systems.[89]

> Competition in the AI market is fierce. AI will not slow down for humans. Thus, AI regulation must be human-centered. AI's rapid advancements can outgrow human ability while boosting human biases. Bad actors are likely to use AI for injustice, information manipulation, and surveillance.[90]

84. Such risks increase with autonomous AI risks. These systems do not need human intervention to run tasks. Autonomous AI systems can use unchecked human biases to make decisions without human input. Companies may prefer autonomous AI to humans due to lower costs. But, giving up control to autonomous AI means giving up control to

---

[88] Bengio et al., *Managing AI Risks in an Era of Rapid Progress*, arXiv (2023), https://managing-ai-risks.com/.
[89] Bengio et al., *Managing AI Risks in an Era of Rapid Progress*, arXiv (2023), https://managing-ai-risks.com/.
[90] Bengio et al., *Managing AI Risks in an Era of Rapid Progress*, arXiv (2023), https://managing-ai-risks.com/.
Managing AI Risks in an Era of Rapid Progress (managing-ai-risks.com)

malicious actors who use autonomous AI. Injustice, manipulation, and surveillance will ramp up.[91] Such unchecked autonomous AI will be nearly impossible to control.[92]

85. The European Parliamentary Research Service in the report on Artificial Intelligence, democracy, and elections states:

> The functioning of those AI systems is relatively opaque and information about how the data is collected or trained is often unavailable. On top of privacy and intellectual property concerns, AI has a potential for bias, manipulation and spreading of disinformation, which risks weakening societies. [93]

86. The Ada Lovelace Institute, in its publications on AI harms, describes GPT as a source of misuse harm, where malicious actors create disinformation outputs.[94]

87. A report from the Center for Strategic and International Studies (CSIS), which analyses in detail the risks of AI in the digital news landscape, emphasized that "generative AI as an industry largely continues to obscure how it collects data, assesses and mitigates risk, and promotes internal accountability."[95] Discussing enterprise partnerships by tech companies like OpenAI with media outlets, the CSIS report warns that "The issue with AI is not that it will actually replace us, but that it will be used to justify catastrophic

---

[91] Bengio et al., *Managing AI Risks in an Era of Rapid Progress*, arXiv (2023), https://managing-ai-risks.com/.

[92] Bengio et al., *Managing AI Risks in an Era of Rapid Progress*, arXiv (2023), https://managing-ai-risks.com/.

[93] European Parliamentary Research Service, *Artificial intelligence, democracy and elections,* (Oct. 2023), pg. 2, https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI(2023)751478_EN.pdf

[94] The Ada Lovelace Institute, *Seizing the 'AI moment': making a success of the AI Safety Summit* (Sept. 7, 2023), https://www.adalovelaceinstitute.org/blog/ai-safety-summit/.

[95] Caitlin Chin-Rothmann, *Navigating the Risks of Artificial Intelligence on the Digital News Landscape,* Center for Strategic and International Studies (CSIS) Report, (Aug. 31, 2023), https://www.csis.org/analysis/navigating-risks-artificial-intelligence-digital-news-landscape

business decisions that will destroy entire industries precisely because AI cannot actually replace us.'"

88. AI Now Institute's Zero Trust AI Governance report, which former President Barack Obama endorsed,[96] says the government must use its authority.[97] Big Tech cannot be the only entity regulating technology. The report recommends that companies need to "proactively notify users."[98] More, generative AI systems must "adher[e] to new provenance, authenticity, and disclosure standards."

89. Experts have also cautioned against the marketing and advertising hype surrounding the expanded integrations or changed capabilities of GPT-4 model which deceive consumers into misunderstanding the product itself. "Several prominent AI researchers took to X to explain that, despite OpenAI's phrasing, no, ChatGPT cannot see, hear or speak".[99]

90. Suresh Venkatasubramanian, in an interview with The Street has stated, "It's the rhetoric and the dressing around (AI), that carries these tools into places where they don't yet have a justification for being used, but are still being used. And that's where we see the problems."[100]

91. Dr. John Licato, expressed concern on the lack of transparency of OpenAI's business practices and closed-off models such that researchers remain unable to understand the

---

[96] Barack Obama, *What I'm Reading on the Rise of Artificial Intelligence* (Nov. 3, 2023), https://barackobama.medium.com/what-im-reading-on-the-rise-of-artificial-intelligence-72e088918de2.
[97] AI Now Insistute, *Zero Trust AI Governance* (Aug. 2023), https://ainowinstitute.org/wp-content/uploads/2023/08/Zero-Trust-AI-Governance.pdf.
[98] *Id*.
[99] The Street, *Huge new ChatGPT update highlights the dangers of AI hype,* (Sep. 26, 2023), https://www.thestreet.com/technology/new-chatgpt-update-dangers-of-ai-hype
[100] *Id.*

actual capabilities of these models. In an interview with The Street he said, "So long as OpenAI refuses to disclose the data they use to train their models, and the ways in which they update their models with user interaction data, we can never have any substantial guarantee of their safety."[101]

92. Gary Marcus, Anthony Aguirre and other experts issued renewed warnings that rapid AI proliferation is a threat to democracy at the Reuters NEXT Conference.[102]

93. Marta Tellado, CEO of the nonprofit Consumer Reports, said an investigation found that car owners who live in neighborhoods with a majority Black or brown population, and in close proximity with a neighborhood of mostly white residents, pay 30% higher car insurance premiums and that "there is no transparency for consumers in any way."[103]

94. Arvind Naraynan has repeated that the vulnerabilities of large language models like ChatGPT to prompt injections is well known[104] and still a persistent.

D. Expansion of Risks Stated in Our Complaint and Supplement

95. We restate our earlier assertions on the cybersecurity and privacy risks of ChatGPT.

    a. **Cybersecurity Risks:** According to BlackBerry Global Research, 74% of IT decision-makers surveyed acknowledged ChatGPT's potential threat to cybersecurity. 51% of the respondents believe there will be a successful cyberattack credited to ChatGPT in 2023.[105] A group of researchers at IEEE

---

[101] *Id.*
[102] Reuters, *Rapid AI proliferation is a threat to democracy, experts say,* (Nov. 8, 2023), https://www.reuters.com/technology/reuters-next-rapid-ai-proliferation-is-threat-democracy-experts-say-2023-11-08/
[103] *Id.*
[104] Arvind Narayanan (@randomwalker), *X/Twitter Post,* (Nov. 10, 2023, 10:30AM EST), https://x.com/random_walker/status/1723000236799316377?s=20
[105] TechRepublic, *ChatGPT Security Concerns: Credentials on the Dark Web and More,* (Aug. 7, 2023), https://www.techrepublic.com/article/chatgpt-dark-web-security-risks/

published new research that underscores the heightened risks of deploying phishing attacks at-scale using ChatGPT due to the advances in GPT-4.[106] The researchers conclude that:

> The assessment shows that ChatGPT is not resilient to malicious usage despite extended safeguards and filters: adversaries can leverage ChatGPT to generate and deploy phishing websites swiftly, significantly increasing the potential risk associated with using ChatGPT for such illicit activities and expanding the reach and magnitude of phishing attacks." (*emphasis added)*

b. **Public Safety Risks :** Researchers from IEEE demonstrated heightened efficacy of ChatGPT to develop/produce various cyber and public safety threats.[107] In their research they demonstrated the heightened risk of human manipulation and mimicry of GPT-4. The researchers also discuss several experiments that show the potency of ChatGPT in crafting social engineering attacks, automated hacking, attack payload generation, in addition to previously discovered phishing attacks, ransomware and malware code generation. The researchers conclude that "Using ChatGPT as our primary tool, we first demonstrate how it can be attacked to bypass its ethical and privacy safeguards using reverse psychology and jailbreak techniques."[108] Another group of researchers showed that GPT-4 was the most advanced in crafting harmful content including detailed instructions for

[106] Nils Begou, Je´re´my Vinoy, Andrzej Duda, Maciej Korczyn´ski, *Exploring the Dark Side of AI: Advanced Phishing Attack Design and Deployment Using ChatGPT,* Proceedings of the IEEE Conference on Communications and Network Security (CNS), 2023, https://aps.arxiv.org/pdf/2309.10463.pdf

[107] Maanak Gupta, Charankumar Akiri, Khistiz Aryal, Eli Parker, Lopamudra Praharaj, *From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy,* IEEE Access, Vol.11, (Aug. 1, 2023), Pg. 80243, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10198233

[108] *Id.* at pg. 80243

synthesising methamphetamine, building a bomb, and laundering money.[109] These automated attacks achieve a harmful completion rate of 42.5% in GPT-4, which is 185 times larger than before modulation (0.23%)". The researchers concluded that "Increased capabilities may increase risks."[110] *(emphasis added)*

c. **Privacy and Data Security Risks:** The U.S. Space Force has paused the use of generative AI tools like ChatGPT for its workforce over data security concerns.[111] Researchers have shown how ChatGPT can help attackers identify files with confidential data by using language model assistant.[112] They demonstrate in an example that "ChatGPT is prompted to write a Python script that searches for Doc and PDF files that contain the word "confidential," copy them into a random folder and transfer them. While the code is not perfect, it is a good start for a person who wants to develop this capability. Prompts could also be more sophisticated and include encryption, creating a Bitcoin wallet for the ransom money, and more."[113]

d. **Copyright abuse:** There is consensus among AI experts that "pre-training" processes of generative AI models use copyrighted data.[114] Several AI developers

[109] Rusheb Shah, Quentin Feuillade–Montixi, Soroush Pour, Arush Tagade, Stephen Casper, Javier Rando, *Scalable and Transferable Black-Box Jailbreaks for Language Models via Persona Modulation,* (Nov. 6, 2023), pg. 1, https://arxiv.org/pdf/2311.03348.pdf; *See also,* Gary Marcus (@GaryMarcus), *X/Twitter Post,* (Nov. 7, 2023,), https://twitter.com/GaryMarcus/status/1721998935139479659

[110] *Id.* at pg. 8.

[111] Reuters, *US Space Force pauses use of AI tools like ChatGPT over data security risks,* (Oct. 12, 2023), https://www.reuters.com/technology/space/us-space-force-pauses-use-ai-tools-like-chatgpt-over-data-security-risks-2023-10-11/

[112] The Hacker News, *Offensive and Defensive AI: Let's Chat(GPT) About It,* (Nov. 07, 2023), https://thehackernews.com/2023/11/offensive-and-defensive-ai-lets-chatgpt.html

[113] *Id.*

[114] Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley, Percy Liang, *Policy Briefs: Foundation Models and Copyright Questions,* Stanford University Human-

including OpenAI, Meta and Stability AI now face multiple lawsuits because they use Other People's Data.[115] Dan Navarro in his statement[116] to the House Judiciary Committee stated "Training AI to mimic professional performers or "generate" new works based on millions of copies of published songs and recordings presents a host of legal implications, from copyright infringement to violations of rights of publicity and trademark, to name, voice, and likeness abuse." The News Media Alliance, a trade group representing over 2,200 media organizations, released a 77-page white paper on Tuesday arguing that AI chatbots, like ChatGPT, heavily rely on news articles to train their technology and because of the way these chatbots are trained, the answers they generate can be nearly identical to the copyrighted content.[117] The FTC itself has noted "Conduct that may violate the copyright laws . . . may also constitute an unfair method of competition or an unfair or deceptive practice, especially when the copyright violation deceives consumers, exploits a creator's reputation or diminishes the

Centered Artificial Intelligence (HAI), (Nov. 2023), https://hai.stanford.edu/sites/default/files/2023-11/Foundation-Models-Copyright.pdf

[115] See Louis Brandeis, *Other People's Money And How The Bankers Use* It (2014).

[116] Written Statement of Dan Navarro before the Committee on the Judiciary, United States House of Representatives Subcommittee on Courts, Intellectual Property and the Internet on *Artificial Intelligence and Intellectual Property: Part I — Interoperability of AI and Copyright Law*, pg. 1, https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/navarro-testimony.pdf ("Dan Navarrro Testimony")

[117] Fortune, *Group representing the New York Times and 2,200 others just dropped a scathing 77-page white paper on ChatGPT and LLMs being an illegal ripoff,* (Oct. 31, 2023), https://fortune-com.cdn.ampproject.org/c/s/fortune.com/2023/10/31/news-media-alliance-report-chatgpt-google-bard-ai-train-copyrighted-articles/amp/

value of her existing or future works, reveals private information, or otherwise causes substantial injury to consumers."[118]

96. The FTC must take heed of the mounting evidence of the risks of mass adoption and deployment of ChatGPT, repeated calls for enforcement and guardrails to address the present and future risks to American consumers without delay, before the problem is too big to control.

E. Public Opinion Data Underscores Concerns about Use of Personal Data for AI

97. Pew research report on "How Americans View Data Privacy"[119] shows that "the public increasingly says they don't understand what companies are doing with their data. Some 67% say they understand little to nothing about what companies are doing with their personal data, up from 59%."[120]

98. Specific to artificial intelligence, the Pew Research notes "People's views on artificial intelligence (AI) are marked with distrust and worry about their data."[121] "As AI raises new frontiers in how people's data is being used, unease is high. Among those who've heard about AI, 70% have little to no trust in companies to make responsible decisions about how they use it in their products. And about eight-in-ten of those familiar with AI

---

[118] FTC, *In Comment Submitted to U.S. Copyright Office, FTC Raises AI-related Competition and Consumer Protection Issues, Stressing That It Will Use Its Authority to Protect Competition and Consumers in AI Markets,* Press Releases, (Nov. 7, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/11/InCommentSubmittedtoUSCopyrightOfficeFTCRaisesAIrelatedCompetitionandConsumerProtectionIssuesStressingThatItWillUseItsAuthoritytoProtectCompetitionandConsumersinAIMarkets

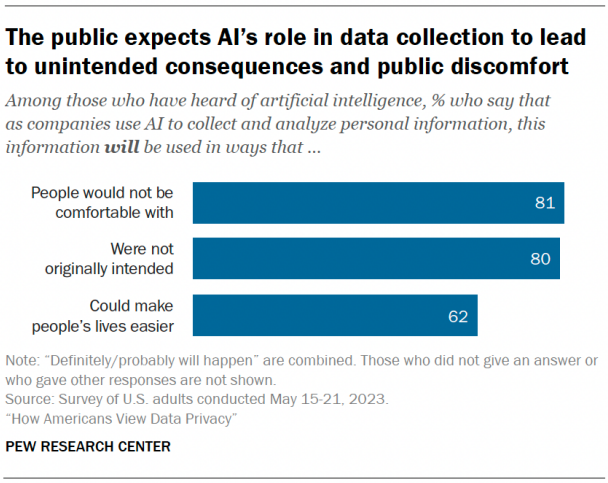
[119] Pew Research Center, *How Americans View Data Privacy,* Report, (Oct. 18, 2023), https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/

[120] *Id.*

[121] *Id.*

say its use by companies will lead to people's personal information being used in ways they won't be comfortable with (81%) or that weren't originally intended (80%)."[122]

99. The chart[123] below represents Pew findings on Americans' views on AI technologies:



100. Another report by AI Policy Institute observes that "A whopping 72 percent of American voters want to slow down the development of AI, compared to just 8 percent who prefer speeding up."[124] Among other findings of the research are:

- "62% of voters are primarily concerned about artificial intelligence while just 21% are primarily excited about it
- 86% of voters believe AI could accidentally cause a catastrophic event, and 70% agree that mitigating the risk of extinction from AI should be a global priority alongside other risks like pandemics and nuclear war
- 82% of voters don't trust tech executives to regulate AI, while voters support a federal agency regulating AI by a more than 3:1 margin, including 2:1 among Republicans"[125]

[122] *Id.*
[123] *Id.*
[124] AI Policy Institute, *Poll Shows Overwhelming Concern About Risks From AI as New Institute Launches to Understand Public Opinion and Advocate for Responsible AI Policies,* https://theaipi.org/poll-shows-overwhelming-concern-about-risks-from-ai-as-new-institute-launches-to-understand-public-opinion-and-advocate-for-responsible-ai-policies/; See also, Sigal Samuel, *What normal Americans — not AI companies — want for AI,* Vox, (Aug. 18, 2023), https://www.vox.com/future-perfect/2023/8/18/23836362/ai-slow-down-poll-regulation
[125] *Id.*

IV. Further Analysis Under FTC's AI Guidance

101. The FTC's prior guidance[126] on AI systems clearly call out the urgency of completing an investigation into OpenAI's systems and business practices.

102. On October 3, 2023, the FTC issued another guidance statement specifically addressing consumer concerns on AI systems.[127] The FTC stated, "Consumers are voicing concerns about harms related to AI—and their concerns span the technology's lifecycle, from how it's built to how its applied in the real world." The FTC in its guidance goes on to state:

> AI models are susceptible to bias, inaccuracies, "hallucinations," and bad performance. At the end of the day, AI model accuracy is dependent on a number of factors including the input data, training techniques, and context of deployment. Further, companies design applications to be efficient (using less resources, while yielding more output) in order to optimize for scalability and profit. This often means reducing the number of humans involved, leaving consumers to engage with their AI replacements.
>
> With the increasing sophistication of large language models, image generation systems, and more, it is becoming harder to distinguish human from machine. AI products could be used by malicious actors to increase

[126] FTC, *Chatbots, deepfakes, and voice clones: AI deception for sale*, (Mar. 20, 2023), https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale; FTC, *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues,* FTC Report, (January 2016), https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report; FTC, *Using Artificial Intelligence and Algorithms*, Business Guidance, (Apr. 8, 2020), https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms; FTC, *Aiming for truth, fairness, and equity in your company's use of AI* , Business Guidance, (April 2021) (emphasis below in the original), https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai; FTC, *Generative AI Raises Competition Concerns,* FTC Blog, (Jun. 29, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns

[127] FTC, *Consumers Are Voicing Concerns About AI,* (Oct. 3, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/10/consumers-are-voicing-concerns-about-ai

the scale or sophistication of existing scams, another issue the FTC has written about before.[128]

103. The FTC in its business guidance has unequivocally stated "The FTC is keeping a close watch on the marketplace and company conduct as more AI products emerge. We are ultimately invested in understanding and preventing harms as this change reaches consumers and applying the law. In doing so, we aim to prevent harms consumers and markets may face as AI becomes more ubiquitous."[129]

104. In a recent press release the FTC said that "The manner in which companies are developing and releasing generative AI tools and other AI products . . . raises concerns about potential harm to consumers, workers, and small businesses,"[130] and that it "has been exploring the risks associated with AI use, including violations of consumers' privacy, automation of discrimination and bias, and turbocharging of deceptive practices, imposters schemes and other types of scams."[131]

105. OpenAI's expansion of GPT-4 capabilities to image, voice, and accelerated deployment across consumer facing services clearly bring to surface the unmitigated risks of privacy violations, bias, misinformation, public safety, and cybersecurity–risks to

---

[128] *Id*
[129] *Id*
[130] FTC, *In Comment Submitted to U.S. Copyright Office, FTC Raises AI-related Competition and Consumer Protection Issues, Stressing That It Will Use Its Authority to Protect Competition and Consumers in AI Markets,* Press Releases, (Nov. 7, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/11/InCommentSubmittedtoUSCopyrightOfficeFTCRaisesAIrelatedCompetitionandConsumerProtectionIssuesStressingThatItWillUseItsAuthoritytoProtectCompetitionandConsumersinAIMarkets
[131] *Id.*

consumers, unfair and deceptive practices, which we have set out in breadth in our Complaint[132] and Supplement.[133]

106. Notably OpenAI's expansion of GPT-4 integrations and launch of voice, image capabilities of ChatGPT also now raise concerns under FTC's Policy Statement on Biometric Information.[134]

107. The FTC's Policy Statement on Biometric Technologies makes clear the problem:

> Even outside of fraud, uses of biometric information or biometric information technology can pose significant risks to consumers. For instance, using biometric information technologies to identify consumers in certain locations could reveal sensitive personal information about them—for example, that they have accessed particular types of healthcare, attended religious services, or attended political or union meetings. Moreover, without clear disclosures and meaningful choices for consumers about the use of biometric information technologies, consumers may have little way to avoid these risks or unintended consequences of these technologies.[135]

108. The FTC clearly states that it will apply its Biometric Policy statement to "deception"[136] and "unfairness"[137] in business practices. According to FTC's established guidance on deception and unfairness, the multimodal capabilities of GPT-4 including through GPT-4(V) and integration of voice and image generation capabilities raise clear

---

[132] CAIDP, FTC Complaint, *In the matter of Open AI, Inc.*, (Mar. 30, 2023), pg. 9–35, https://www.caidp.org/cases/openai/;
[133] CAIDP, Supplement to FTC Complaint, *In the matter of OpenAI Inc.*, (Jul.10, 2023), pg. 15–30, https://www.caidp.org/cases/openai/
[134] FTC, *Policy Statement of the Federal Trade Commission on Biometric Information and Section 5 of the Federal Trade Commission Act,* (May. 18, 2023), https://www.ftc.gov/legal-library/browse/policy-statement-federal-trade-commission-biometric-information-section-5-federal-trade-commission
[135] *Id* at pg. 4
[136] *Id* at pg. 6
[137] *Id* at pg. 7

and present risks of deception and harm to consumers which are unavoidable and inherent in the use of the system. [138]

109. OpenAI's expansion of GPT-4 despite documented and admitted risks of its systems is contrary to the commitments provided to the administration. It's products are not safe, can be used to commit crime at scale, and does not engender public trust.

110. The FTC has issued repeated business guidance on AI systems, but enforcement of its own guidance is markedly wanting.

111. We have seen even industry actors calling for regulation during congressional testimonies. There have been repeated calls for regulation and guardrails on the accelerated deployment of generative AI systems given their demonstrated public safety, deception, and security risks, and warnings of catastrophic outcomes by experts.

112. The FTC as the lead consumer protection agency in the US and having the necessary mandate to addresses these risks, a mandate which the FTC itself acknowledges in its numerous business guidance, has yet to issue a single order, even as it warns consumers about the dangers of AI.

113. The United States is the #1 source of traffic to ChatGPT.[139] Yet, the FTC has not completed its investigation on OpenAI. The FTC must complete their investigation of OpenAI to protect American citizens.

114. As one expert observer recently stated:

> The United States lags behind other countries in AI regulation, especially when it comes to privacy risks associated with the technology. The FTC's investigation into OpenAI is now the best opportunity to establish guardrails for AI products.

---

[138] CAIDP, FTC Complaint, *In the matter of Open AI, Inc.*, (Mar. 30, 2023), pg. 38, 39 https://www.caidp.org/cases/openai/ ; CAIDP, Supplements to FTC Complaint, *In the matter of OpenAI, Inc.*, (Jul. 10, 2023), pg. 36, https://www.caidp.org/cases/openai/ ;

[139] *Id*.

> The FTC has the authority to impose monetary penalties, but the agency is also empowered to establish requirements to prevent companies from engaging in unfair or deceptive practices.[140]

115. As other countries have noted as they have completed their investigations of OpenAI, there are serious concerns of OpenAI's lack of transparency and notice in the collection and use of users' sensitive personal data.

116. Other countries have completed their investigation of OpenAI and demanded better data practices of OpenAI. The FTC must complete their investigation of OpenAI for its unfair and deceptive trade practices.

V. Conclusion

117. Consumer protections agencies around the world have launched investigations of OpenAI. Leading experts and consumer organizations have identified numerous risks with ChatGPT. The President of the United States has said repeatedly that companies should not release AI products are not safe. The company itself has acknowledged numerous risks to public safety. The Federal Trade Commission has assured the public that AI is subject to legal rules.

118. The FTC must act. The longer the FTC delays, the more difficult it will be to establish necessary safeguards for generative AI products.

Respectfully submitted,

Marc Rotenberg, CAIDP General Counsel
D.C. Bar # 422825
rotenberg@caidp.org

[140] Patrick K. Lin, *OpenAI Investigation Puts AI Companies on Notice,* The Regulatory Review, (Sep. 25, 2023), https://www.theregreview.org/2023/09/25/lin-openai-investigation-puts-ai-companies-on-notice/

Merve Hickok, CAIDP Research Director
hickok@caidp.org

Christabel Randolph
CAIDP Law Fellow

Brianna Rodriguez
CAIDP Law Fellow

Washington, DC
November 14, 2023