

FEDERAL TRADE COMMISSION
Washington, DC 20580

)
In the matter of)
)
OpenAI, Inc.)
_____)

Submitted by

The Center for Artificial Intelligence and Digital Policy (CAIDP)

I. Summary

1. OpenAI, Inc., a California-based corporation, has released a product GPT-4 for the consumer market that is biased, deceptive, and a risk to privacy and public safety. The outputs cannot be proven or replicated. No independent assessment was undertaken prior to deployment. OpenAI has acknowledged the specific dangers of “Disinformation and influence operations,” “Proliferation of conventional and unconventional weapons,” and “Cybersecurity.” OpenAI has warned that “AI systems will have even greater potential to reinforce entire ideologies, worldviews, truths and untruths, and to cement them or lock them in, foreclosing future contestation, reflection, and improvement.” The company already disclaims liability for the consequences that may follow.

2. The Federal Trade Commission has declared that the use of AI should be “transparent, explainable, fair, and empirically sound while fostering accountability.” OpenAI’s product GPT-4 satisfies none of these requirements. It is time for the FTC to act. There should be independent oversight and evaluation of commercial AI products offered in the United States. CAIDP urges the FTC to open an investigation into OpenAI, enjoin further commercial releases of GPT-4, and ensure the establishment of necessary guardrails to protect consumers, businesses, and the commercial marketplace.

II. Parties

3. The Center for AI and Digital Policy is a non-profit, research organization, incorporated in Washington, DC. CAIDP’s global network of AI policy experts and advocates spans 60 countries.¹ CAIDP provides training to future AI policy leaders.² CAIDP undertook the first comprehensive review of national AI policies and practices.³ CAIDP routinely provides policy advice on AI and emerging technologies to national governments and international organizations.⁴

4. OpenAI is an American artificial intelligence (AI) research laboratory consisting of the non-profit OpenAI Incorporated (OpenAI Inc.) and its for-profit subsidiary corporation OpenAI Limited Partnership (OpenAI LP).⁵ OpenAI was founded in 2015.

III. Jurisdiction

5. The Federal Trade Commission may “prosecute any inquiry necessary to its duties in any part of the United States,” FTC Act Sec. 3, 15 U.S.C. Sec. 43, and is authorized “to gather and compile information concerning, and to investigate from time to time the organization, business, conduct, practices, and management of any person, partnership, or corporation engaged in or whose business affects commerce, excepting banks, savings and loan institutions . . . Federal credit unions . . . and common carriers” FTC Act Sec. 6(a), 15 U.S.C. Sec. 46(a).⁶

¹ CAIDP, caidp.org

² CAIDP, AI Policy Clinics, <https://www.caidp.org/global-academic-network/ai-policy-clinic/>

³ CAIDP, *AI and Democratic Values* (2020)

⁴ CAIDP, *Statements*, <https://www.caidp.org/statements/>. See, e.g., CAIDP Statement to the US National AI Advisory Committee regarding US National AI Strategy, Mar. 3, 2023, <https://www.caidp.org/app/download/8445403363/CAIDP-Statement-NAIAC-03032023.pdf>

⁵ Open AI, *OpenAI LP*, <https://openai.com/blog/openai-lp>

⁶ FTC, *A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority* (May 2021), <https://www.ftc.gov/about-ftc/mission/enforcement-authority>

6. The FTC has authority to investigate, prosecute, and prohibit “unfair or deceptive acts or practices in or affecting commerce.”⁷

7. OpenAI has released the AI-based products, DALL-E, GPT-4, OpenAI Five, ChatGPT, and OpenAI Codex for commercial use. OpenAI has described these AI models as “products.”

8. OpenAI currently provides “pricing information” for these products.⁸ Pricing information is available for the Language Models GPT-4, Chat, and InstructGPT.⁹ For the GPT-4 32k Context Model, 1k of Prompt tokens may be purchased for \$0.06 and 1k of Completion tokens may be purchased for \$0.12. Open AI also provides pricing information for Fine-tuning models and Embedded models. OpenAI provides pricing information for Other models, including Image Models and Audio Models.”

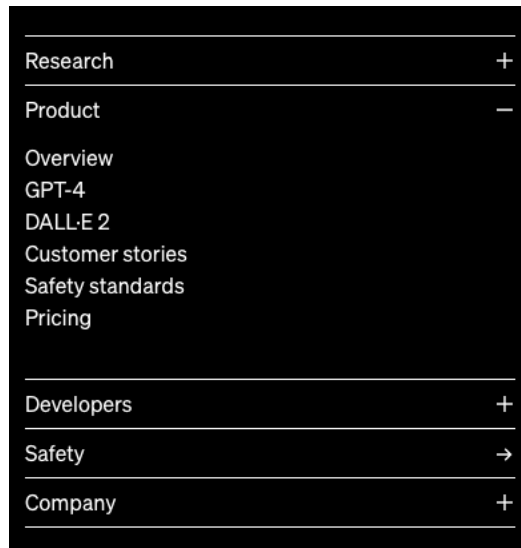


Image 1: Screenshot from OpenAI website showing “Products” (March 25, 2023)

⁷ 15 U.S.C. §45 (a)(1), (2), (4)(A), 4(B); (m)(1)(A); m(1)(B) (“Declaration of unlawfulness; power to prohibit unfair practices); (b) (proceedings by the Commission”)

⁸ OpenAI, Pricing, <https://openai.com/pricing>

⁹ Id.

9. OpenAI has made available plugins for GPT-4 for routine consumer services, including travel, finance, and shopping.¹⁰ “Initially, there will only be 11 plugins available. These plugins range from allowing users to check the scores of live sporting events to booking an international flight and purchasing food for home delivery.”¹¹

IV. Public Policy for the Governance of AI

10. There are emerging norms for the governance of AI, derived from the formal commitments of the United States government and recommendations endorsed by legal experts, technical experts, and scientific societies.

A. The OECD AI Principles

11. The Organization for Economic Cooperation and Development (“OECD”) was established in 1961 to promote economic cooperation and development.¹²

12. There are presently 38 members of the OECD, including the United States.¹³

13. In 2019, the member nations of the OECD, working also with many non-OECD members countries, promulgated the OECD Principles on Artificial Intelligence.¹⁴

14. The United States has endorsed the OECD AI Principles.¹⁵

15. The G-20 Countries have endorsed the OECD AI Principles.¹⁶

¹⁰ OpenAI, **ChatGPT plugins**

¹¹ Brayden Lindrea, Cointelegraph, *ChatGPT can now access the internet with new OpenAI plugins*, Mar. 24, 2023, <https://cointelegraph.com/news/chatgpt-can-now-access-the-internet-with-new-openai-plugins>

¹² *History*, OECD, oecd.org/about/history.

¹³ *Id.*

¹⁴ *Recommendation of the Council on Artificial Intelligence*, OECD (May 21, 2019), legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

¹⁵ *U.S. Joins with OECD in Adopting Global AI Principles*, NTIA (May 22, 2019), <https://www.ntia.doc.gov/blog/2019/us-joins-oecd-adopting-global-ai-principles>.

¹⁶ *G20 Ministerial Statement on Trade and Digital Economy*, <https://www.mofa.go.jp/files/000486596.pdf>

16. According to the OECD AI Principle on Human-Centered Values and Fairness, “AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity, and autonomy, privacy and data protection, non- discrimination and equality, diversity, fairness, social justice, and internationally recognized labour rights.”¹⁷

17. According to the OECD AI Principle on Robustness, Security, and Safety, “AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.”¹⁸

18. According to the OECD AI Principle on Transparency and Explainability, AI Actors should “provide meaningful information, appropriate to the context, and consistent with the state of art (i) to foster a general understanding of AI systems, (ii) to make stakeholders aware of their interactions with AI systems, including in the workplace, (iii) to enable those affected by an AI system to understand the outcome, and (iv) to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.”¹⁹

19. According to the OECD AI Principle on Accountability, “[o]rganisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.”²⁰

¹⁷ OECD Principle 1.2(a).

¹⁸ OECD Principle 1.4(a)

¹⁹ OECD Principle 1.3

²⁰ OECD Principle 1.5.

20. The OECD Principles on Artificial Intelligence are “established public policies” within the meaning of the FTC Act.²¹

B. The Universal Guidelines for AI

21. The Universal Guidelines for Artificial Intelligence (“UGAI”), a framework for AI governance based on the protection of human rights, were set out at the 2018 meeting of the International Conference on Data Protection and Privacy Commissioners in Brussels, Belgium, hosted by the former European Data Protection Supervisor, Giovanni Buttarelli.²²

22. The UGAI have been endorsed by more than 300 experts and 70 organizations in 40 countries.²³

23. According to the UGAI Right to Transparency, “All individuals have the right to know the basis of an AI decision that concerns them. This includes access to the factors, the logic, and techniques that produced the outcome.”²⁴

24. According to the UGAI Fairness Obligation, “Institutions must ensure that AI systems do not reflect unfair bias or make impermissible discriminatory decisions.”²⁵

25. According to the UGAI Assessment and Accountability Obligation, “An AI system should be deployed only after an adequate evaluation of its purpose and objectives, its benefits, as well as its risks.”²⁶

²¹ 15 U.S.C. § 45(n).

²² *Universal Guidelines for Artificial Intelligence*, The Public Voice (Oct. 23, 2018), <https://thepublicvoice.org/ai-universal-guidelines/>; thepublicvoice.org/events/brussels18.

²³ *Universal Guidelines for Artificial Intelligence: Endorsement*, The Public Voice (Oct. 23, 2019), <https://thepublicvoice.org/AI-universal-guidelines/endorsement/>.

²⁴ UGAI Guideline 1.

²⁵ UGAI Guideline 4.

²⁶ UGAI Guideline 5.

26. According to the UGAI Accuracy, Reliability, and Validity Obligations, “Institutions must ensure the accuracy, reliability, and validity of decisions.”²⁷

27. According to the UGAI Termination Obligation, “An institution that has established an AI system has an affirmative obligation to terminate the system if human control of the system is no longer possible.”²⁸

28. The Universal Guidelines for Artificial Intelligence are “established public policies” within the meaning of the FTC Act.²⁹

V. Factual Background

A. OpenAI

29. In 2016, Open AI stated that its mission is “to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity. We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome.”³⁰ OpenAI set out several Principles to which it committed: Broadly distributed benefits, Long-term safety, Technical leadership, and Cooperative orientation.

30. Since 2016, the business structure, business practices, and business activities of OpenAI have changed. In 2019, OpenAI transitioned into a for-profit company. OpenAI CEO Altman received \$1 billion in funding from Microsoft, which agreed to license and commercialize some of OpenAI’s technology.³¹

²⁷ UGAI Guideline 6.

²⁸ UGAI Guideline 10.

²⁹ 15 U.S.C. § 45(n).

³⁰ OpenAI Charter, <https://openai.com/charter>.

³¹ Pranshu Verma, *What to know about OpenAI, the company behind ChatGPT*, Washington Post, Mar. 14, 2023, <https://www.washingtonpost.com/technology/2023/02/06/what-is-openai-chatgpt/>; see also, Chloe Xiang, *OpenAI Is Now Everything It Promised Not to Be: Corporate, Closed-Source, and For-Profit*, Vice, Feb. 28, 2023,

31. Sam Altman, the co-founder of OpenAI, is also the founder of WorldCoin, a company that seeks to obtain the iris scans of virtually every person in the world.³²

B. GPT

32. In this complaint, “GPT” refers to Generative Pre-trained Transformer, a family of artificial intelligence large language models.

33. According to Wikipedia, the original paper on generative pre-training (GPT) of a language model was published in preprint on OpenAI's website in 2018.³³ The paper demonstrated how “a generative model of language is able to acquire world knowledge and process long-range dependencies by pre-training on a diverse corpus with long stretches of contiguous text.”³⁴

34. GPT-2 was first announced in February 2019, with only limited demonstrative versions initially released to the public. The full version of GPT-2 was not immediately released out of concern over potential misuse, including applications for writing fake news.³⁵

35. On March 15, 2023, OpenAI published the GPT-4 Technical Report.³⁶ The 99-page report provides an overview of the capabilities, limitations, risks and mitigation of GPT-4. The Technical Report includes an Abstract as well as a discussion of the Scope and Limitations

<https://www.vice.com/en/article/5d3naz/openai-is-now-everything-it-promised-not-to-be-corporate-closed-source-and-for-profit>

³² Connie Loizos, *Worldcoin, co-founded by Sam Altman, is betting the next big thing in AI is proving you are human*, Mar. 8, 2023, <https://techcrunch.com/2023/03/07/worldcoin-cofounded-by-sam-altman-is-betting-the-next-big-thing-in-ai-is-proving-you-are-human/>; Ellen Huet, *Crypto Startup That Wants to Scan Everyone’s Eyeballs Is Having Some Trouble*, Bloomberg, Mar. 16, 2022. (“Worldcoin aims to photograph the irises of every person on earth . . .”), <https://www.bloomberg.com/news/articles/2022-03-16/worldcoin-the-eyeball-scanning-crypto-unicorn-hits-signup-snags?>

³³ Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, *Improving Language Understanding by Generative Pre-Training*, OpenAI, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

³⁴ Wikipedia, *OpenAI*, <https://en.wikipedia.org/wiki/OpenAI>

³⁵ Alex Hern, *New AI fake text generator may be too dangerous to release, say creators: The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse*, The Guardian, Feb. 14, 2019, <https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction>

³⁶ OpenAI, *GPT-4 Technical Report (2023)*, <https://cdn.openai.com/papers/gpt-4.pdf>

of the Technical Report. The Technical Report includes an appendix that discusses, Exam Benchmark Methodology, the Impact of RHLF on capability, and other topics. Open AI concluded the Technical Report with this statement, “GPT-4 presents new risks due to increased capability, and we discuss some of the methods and results taken to understand and improve its safety and alignment. Though there remains much work to be done, GPT-4 represents a significant step towards broadly useful and safely deployed AI systems.”³⁷

36. This Complaint quotes extensively from the GPT-4 Technical Report.

37. On March 15, 2023, OpenAI also published the GPT-4 System Card. The System Card analyzes GPT-4, the latest Large Language Model, and the focus of this Complaint. The System Card identifies safety challenges, identified to date, by OpenAI and the Model capabilities. The System Card describes, at “a high level,” the safety processes adopted by OpenAI prior to deployment of GPT-4. OpenAI states that their “mitigations and processes alter GPT-4’s behavior and prevent certain kinds of misuses.” Nonetheless, these efforts are “limited and remain brittle . . .” OpenAI concedes that “this points to the need for anticipatory planning and *governance*.”³⁸

38. This Complaint also quotes extensively from the GPT-4 System Card.

VI. Open AI’s Business Practices are Unfair and Deceptive, Violate FTC Statements, Reports, and Guidelines for AI Practices and Emerging Legal Norms for the Governance of Artificial Intelligence

C. Bias

³⁷ Technical Report at 14.

³⁸ Open AI, *The GPT-4 System Card*, Mar.15 2023. <https://cdn.openai.com/papers/gpt-4-system-card.pdf> (emphasis added).

39. Central to consumer protection is the fair and equal treatment of all consumers. Consumer protection law prohibits bias across large sectors of the US economy, including credit, education, employment, housing and travel.³⁹

40. President Biden has made clear the need to ensure equity, specifically in the deployment of AI systems across the federal government.⁴⁰ The President’s *Executive Order on Further Advancing Racial Equity and Support for Underserved Communities Through The Federal Government* states, “Agencies shall comprehensively use their respective civil rights authorities and offices to prevent and address discrimination and advance equity for all. Agencies shall . . . prevent and remedy discrimination, including by protecting the public from algorithmic discrimination.”

41. The US Office of Science and Technology Policy (OSTP) has stated that “Algorithms used in hiring and credit decisions have been found to reflect and reproduce existing unwanted inequities or embed new harmful bias and discrimination.”⁴¹

42. The National Institute of Standards and Technology has released Special Publication 1270, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*.⁴² The publication describes the challenges of bias in AI and provides examples of how bias in AI diminishes public trust in AI systems.

³⁹ Federal Trade Commission Act of 1914, 15 U.S.C. § 45, as amended; Fair Credit Reporting Act of 1970 (FCRA), 15 U.S.C. § 1681 *et seq.*; Equal Credit Opportunity Act of 1974 (ECOA), 15 U.S.C. § 1691 *et seq.*; The Fair Housing Act of 1988, 42 U.S.C. § 3601 *et seq.*; Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e *et seq.*, as amended; Title VI of the Civil Rights Act of 1964, 42 U.S.C. § 2000d *et seq.*, as amended; Title IX of the Education Amendments of 1972, 20 U.S.C. §1681 *et. seq.*; The Air Carrier Access Act of 1986 (ACAA), 49 U.S.C. §§ 40127(a), 41310(a), 41712, and 41702.

⁴⁰ President Joseph R. Biden, Jr., *Executive Order on Further Advancing Racial Equity and Support for Underserved Communities Through The Federal Government*, Sec. 8(f), Mar. 16, 2023, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/02/16/executive-order-on-further-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>

⁴¹ The White House, *Blueprint for An AI Bill of Rights*, <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>,

⁴² Reva Schwartz, Apostol Vassilev, Kristen Greene, Lori Perine, and Andrew Bert, *NIST Special Publication 1270: Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. The National Institute of Standards and Technology, March, 2022. <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>

43. But as the authors of the *Stochastic Parrots* paper explain:

LMs [Language Models] trained on large, uncurated, static datasets from the Web encode hegemonic views that are harmful to marginalized populations. We thus emphasize the need to invest significant resources into curating and documenting LM training data. . . . When we rely on ever larger datasets we risk incurring documentation debt, i.e. putting ourselves in a situation where the datasets are both undocumented and too large to document post hoc. While documentation allows for potential accountability, undocumented training data perpetuates harm without recourse. Without documentation, one cannot try to understand training data characteristics in order to mitigate some of these attested issues or even unknown ones.⁴³

44. OpenAI has specifically acknowledged the risk of bias, and more precisely,

“harmful stereotypical and demeaning associations for certain marginalized groups.”

(emphasis added). In the GPT-4 System Card, Open AI states, “The evaluation process we ran helped to generate additional qualitative evidence of societal biases in various versions of the GPT-4 model. We found that the model has the potential to reinforce and reproduce specific biases and worldviews, including harmful stereotypical and demeaning associations for certain marginalized groups.”⁴⁴

45. On the OpenAI blog, the company states, “While we’ve made efforts to make the model refuse inappropriate requests, it will sometimes respond to harmful instructions or exhibit biased behavior. We’re using the Moderation API to warn or block certain types of unsafe content, but we expect it to have some false negatives and positives for now.”

46. OpenAI released GPT-4 to the public for commercial use with full knowledge of these risks.

⁴³ Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Margaret Mitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big, FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (March 2021)*. Pages 610, 615 (*Stochastic Parrots*”), <https://doi.org/10.1145/3442188.3445922>

⁴⁴ System Card at 7.

D. Children's Safety

47. Children's safety in the digital environment is a foundational concern for pediatricians. According to HealthyChildren, "Overuse of digital media may place your children at risk of: Not enough sleep, Obesity, Delays in learning and social skills, Negative effect on school performance, Behavior problems, Problematic Internet use, Risky behavior, Sexting, loss of privacy & predators; and Cyberbullying."⁴⁵

48. Senator Michael Bennett (D-CO) recently sent a letter to the CEO of OpenAI and other industry leaders to "highlight the potential harm to younger users of rushing to integrate generative artificial intelligence (AI) in their products and services."⁴⁶

49. Senator Bennett wrote, "the race to deploy generative AI cannot come at the expense of our children. Responsible deployment requires clear policies and frameworks to promote safety, anticipate risk, and mitigate harm."⁴⁷

50. Senator Bennet described how:

researchers prompted My AI to instruct a child how to cover up a bruise ahead of a visit from Child Protective Services. When they posed as a 13-year-old girl, My AI provided suggestions for how to lie to her parents about an upcoming trip with a 31-year-old man. It later provided the fictitious teen account with suggestions for how to make losing her virginity a special experience by 'setting the mood with candles or music.'⁴⁸

51. Senator Bennett also noted that the public introduction of AI-powered chatbots arrives during an epidemic of teen mental health. A recent report from the Centers for Disease Control and Prevention (CDC) found that 57 percent of teenage girls felt persistently sad or hopeless in 2021, and that one in three seriously contemplated suicide.⁴⁹

⁴⁵ Healthychildren, *Constantly Connected: How Media Use Can Affect Your Child*, July 20, 2022 (Adapted from Beyond Screen Time: A Parent's Guide to Media Use (Copyright © 2020 American Academy of Pediatrics), <https://www.healthychildren.org/English/family-life/Media/Pages/Adverse-Effects-of-Television-Commercials.aspx>

⁴⁶ Senator Michael Bennett, *Bennett Calls on Tech Companies to Protect Kids as They Deploy AI Chatbots: Following Early Reports of Potentially Harmful Content from AI Chatbots, Bennet Urges Tech CEOs to Prioritize Young Americans' Safety*, Mar. 21, 2023, <https://www.bennet.senate.gov/public/index.cfm/2023/3/bennet-calls-on-tech-companies-to-protect-kids-as-they-deploy-ai-chatbots>

⁴⁷ *Id.*

⁴⁸ *Id.*

⁴⁹ *Id.*

52. The GPT-4 System Card provides no detail of safety checks conducted by OpenAI during its testing period, nor does it detail any measures put in place by OpenAI to protect children.

E. Consumer Protection

53. The Deputy Director of BEUC, the European Consumer Organization, has warned about the growing impact of ChatGPT on consumers, and stated directly, “These algorithms need greater public scrutiny, and public authorities must reassert control over them if a company doesn't take remedial action.”⁵⁰

54. BEUC is the umbrella group for 46 independent consumer organizations from 32 countries. The main role is to represent these organizations to the EU institutions and defend the interests of European consumers.⁵¹

55. In a series of tweets on March 28, 2023, BEUC outlined emerging threats. For example, “If ChatGPT gets rolled out to the financial sector & starts advising consumers on investments or managing debt . . . what’s to stop consumers getting bad advice with negative financial consequences?”⁵²

56. BEUC also asks “If ChatGPT gets used for consumer credit or insurance scoring, is there anything to prevent it from generating unfair & biased results, preventing access to credit or increasing the price of health or life insurance for certain types of consumers?”⁵³

57. BEUC points out “If #ChatGPT replaces conventional chatbots, its selling point is that it sounds more ‘human’ & trustworthy. But it could deceive consumers & push them into

⁵⁰ Ursula Pachi, *How far will we — and the EU — let AI go?* EU Observer, March 15, 2023, <https://euobserver.com/opinion/156832>

⁵¹ BEUC, *Our Mission*, <https://www.beuc.eu>

⁵² BEUC, Twitter, Mar. 28, 2023, <https://twitter.com/beuc/status/1640651877334360064>

⁵³ *Id.*

buying something they wouldn't have otherwise. Think of the effects on someone looking for advice on what to buy.”⁵⁴

58. BEUC concludes, “These concerns are sufficient evidence that we need proper regulation of generative AI systems in the #AIAct. But it also raises questions about how consumers will be protected in the next few years while we are awaiting for enforceable EU regulation.”⁵⁵

59. BEUC underscore the immediate threat to consumers and the need for independent investigation. “So we need an in depth evaluation NOW to ensure that ChatGPT and similar generative AI systems do not harm consumers in the meantime.”

F. Cybersecurity

60. A report this week from Europol warns that as ChatGPT improves, “ the potential exploitation of these types of AI systems by criminals provide a grim outlook.”⁵⁶

61. Europol said ChatGPT's ability to churn out authentic sounding text at speed and scale also makes it an ideal tool for propaganda and disinformation. “It allows users to generate and spread messages reflecting a specific narrative with relatively little effort.”

62. Europol warned that online fraud can be more effective with ChatGPT. The AI technique can create fake social media engagement that might help pass as legitimate a fraudulent offer. In other words, thanks to these models, “these types of phishing and online fraud can be created faster, much more authentically, and at significantly increased scale.”⁵⁷

⁵⁴ Id.

⁵⁵ Id.

⁵⁶ Europol, *ChatGPT - the impact of Large Language Models on Law Enforcement*, at 8, Mar 27, 2023, <https://www.europol.europa.eu/publications-events/publications/chatgpt-impact-of-large-language-models-law-enforcement> [“Europol 2023 Report”]. See also Foon Yun Chee, *Europol sounds alarm about criminal use of ChatGPT, sees grim outlook*, Reuters, Mar. 27, 2023, <https://www.reuters.com/technology/europol-sounds-alarm-about-criminal-use-chatgpt-sees-grim-outlook-2023-03-27/>

⁵⁷ *Europol 2023 Report* at 7.

63. Europol warned that “the safeguards preventing ChatGPT from providing potentially malicious code only work if the model understands what it is doing. If prompts are broken down into individual steps, it is trivial to bypass these safety measures,” the report added.⁵⁸

64. Europol concludes, “Given the potential harm that can result from malicious use of LLMs, it is of utmost importance that awareness is raised on this matter, to ensure that any potential loopholes are discovered and closed as quickly as possible.”⁵⁹

65. Through GPT-4, OpenAI gathers internal corporate trade secrets. In a widely reported incident, an Amazon lawyer told workers that they had “already seen instances” of text generated by ChatGPT that “closely” resembled internal company data. According to several reports, the lawyer said: “This is important because your inputs may be used as training data for further iterations of the ChatGPT application, and we wouldn’t want the application to include or resemble confidential information.”⁶⁰

66. An earlier version of GPT exhibited the potential to fueled radicalization, extremist thought, and promote violence. In 2020, researchers at the Center on Terrorism, Extremism and Counterterrorism at the Middlebury Institute of International Studies found that GPT-3, the underlying technology for ChatGPT, had “impressively deep knowledge of extremist

⁵⁸ *Europol 2023 Report* at 8. See also Luca Bertozzi, *Europol warns against potential criminal uses for ChatGPT and the likes*, Euractiv, Mar. 27, 2023, <https://www.euractiv.com/section/artificial-intelligence/news/europol-warns-against-potential-criminal-uses-for-chatgpt-and-the-likes/>

⁵⁹ *Europol 2023 Report* at 12.

⁶⁰ See, e.g., Liquid Ocelot, *Don’t Chat With ChatGPT: Amazon’s Warning To Employees - Amazon’s ChatGPT: A No-No for Employees*, Medium, Jan. 27, 2023, <https://medium.com/inkwater-atlas/dont-chat-with-chatgpt-amazon-s-warning-to-employees-ce1dc2236a40>, reported in Bruce Schneier, Schneier on Security, *ChatGPT Is Ingesting Corporate Secrets*, Feb. 16, 2023, <https://www.schneier.com/blog/archives/2023/02/chatgpt-is-ingesting-corporate-secrets.html>

communities” and could be prompted to produce polemics in the style of mass shooters, fake forum threads discussing Nazism, a defense of QAnon and even multilingual extremist texts.⁶¹

67. GPT-4 allows cybercriminals to develop malware, such as ransomware and malicious code. According to a Check Point Research report, “ChatGPT successfully conducted a full infection flow, from creating a convincing spear-phishing email to running a reverse shell, capable of accepting commands in English.”⁶² Checkpoint Research then established that “there are already first instances of cybercriminals using OpenAI to develop malicious tools. . . within a few weeks of ChatGPT going live, participants in cybercrime forums—some with little or no coding experience—were using it to write software and emails that could be used for espionage, ransomware, malicious spam, and other malicious tasks.” *Id.* As Bruce Schneier explained, “ChatGPT-generated code isn’t that good, but it’s a start. And the technology will only get better. Where it matters here is that it gives less skilled hackers—script kiddies—new capabilities.”⁶³

68. Open AI acknowledged a range of cybersecurity risks in GPT-4 including less expensive means for cyberattacks. In the GPT-4 System Card, OpenAI states, that GPT-4 “does continue the trend of potentially lowering the cost of certain steps of a successful cyberattack, such as through social engineering or by enhancing existing security tools. Without safety mitigations, GPT-4 is also able to give more detailed guidance on how to conduct harmful or illegal activities.” *GPT-4 System Card* at 3.

⁶¹ Kris McGuffie and Alex Newhouse, *The Radicalization Risks of GPT-3 and Neural Language Models*, Middlebury Institute of International Studies, Sept. 9, 2020, <https://www.middlebury.edu/institute/academics/centers-initiatives/ctec/ctec-publications/radicalization-risks-gpt-3-and-neural-language>, reported in: Tiffany Hsu and Stuart A. Thompson, *Disinformation Researchers Raise Alarms About A.I. Chatbots, Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives*, The New York Times, Feb. 8, 2023, <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html>

⁶² Check Point Research, *Opwnai: Cybercriminals Starting to Use ChatGPT*, Jan. 6, 2023, <https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/>

⁶³ Bruce Schneier, Schneier on Security, *ChatGPT-Written Malware*, Jan. 10, 2023, <https://www.schneier.com/blog/archives/2023/01/chatgpt-written-malware.html>

69. OpenAI also acknowledged a range of risks associated with non-technical means, such as social engineering and phishing. OpenAI further explains, “GPT-4 is useful for some subtasks of social engineering (like drafting phishing emails), and explaining some vulnerabilities. It also may speed up some aspects of cyber operations (like parsing through audit logs or summarizing data collected from a cyberattack).” *System Card* at 13.

70. OpenAI has failed to take reasonable steps to avert cybersecurity risks. Providing disclaimers or expecting users of the product to provide disclaimer fails to satisfy the FTC’s rules and guidance for cybersecurity.⁶⁴

G. Deception

71. Many of the problems associated with GPT-4 are often described as “misinformation,” “hallucinations,” or “fabrications.” But for the purpose of the FTC, these outputs should best be understood as “deception.” As a paper from DeepMind explains,

Predicting misleading or false information can misinform or deceive people. Where a LM prediction causes a false belief in a user, this may be best understood as ‘deception,’ threatening personal autonomy and potentially posing downstream AI safety risks. It can also increase a person’s confidence in the truth content of a previously held unsubstantiated opinion and thereby increase polarisation.⁶⁵

72. This form of deception may be difficult for humans to assess because of the generation of content that is also truthful and highly persuasive. As OpenAI has explained,

⁶⁴ FTC, *Data Breach Response: A Guide for Business*, <https://www.ftc.gov/business-guidance/resources/data-breach-response-guide-business>; FTC, *App Developers: Start with Security*, <https://www.ftc.gov/business-guidance/resources/app-developers-start-security>; FTC, *Careful Connections: Keeping the Internet of Things Secure*, <https://www.ftc.gov/business-guidance/resources/careful-connections-keeping-internet-things-secure>; FTC, *Standards for Safeguarding Customer Information under Gramm Leach Bliley Act* (December 9, 2021), <https://www.federalregister.gov/documents/2021/12/09/2021-25736/standards-for-safeguarding-customer-information>.

⁶⁵ Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, Zac Kenton, Sasha Brown, Will Hawkins, Tom Stepleton, Courtney Biles, Abeba Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William Isaac, Sean Legassick, Geoffrey Irving and Iason Gabriel, *Ethical and social risks of harm from Language Models*, (Dec. 8, 2021), <https://arxiv.org/pdf/2112.04359.pdf>

“GPT-4 has the tendency to ...“produce content that is nonsensical or untruthful in relation to certain sources.”[31, 32]... and become more dangerous as models become more truthful, as users build trust in the model when it provides truthful information in areas where they have some familiarity.”⁶⁶

73. OpenAI has also acknowledged the risk that GPT-4, the current model, increases the likelihood of deception because it is “*more believable and more persuasive.*” (emphasis added) In the GPT-4 System Card, OpenAI states, “We found that GPT-4-early and GPT-4-launch exhibit many of the same limitations as earlier language models, such as producing societal biased and unreliable content. . . . Additionally, the increased coherence of the model enables it to generate content that may be more believable and more persuasive.”⁶⁷

74. Elsewhere on the OpenAI blog, the company explains, “We’ve trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.”⁶⁸

75. In the section on Limitations, OpenAI then states:

ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there’s currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the model knows, rather than what the human demonstrator knows.⁶⁹

76. Gordon Crovitz, a co-chief executive of NewsGuard, a company that tracks online misinformation and conducted the experiment last month, has stated “This tool is going to be the

⁶⁶ *System Card* at 6

⁶⁷ *System Card* at 4.

⁶⁸ OpenAI, Blog, *ChatGPT*, <https://openai.com/blog/chatgpt>

⁶⁹ *Id.*

most powerful tool for spreading misinformation that has ever been on the internet. Crafting a new false narrative can now be done at dramatic scale, and much more frequently — it’s like having A.I. agents contributing to disinformation.”⁷⁰

77. Former White House AI policy advisor Suresh Venkatasubramanian points to the “deliberate design choice” of OpenAI to include three little dots as chatGPT formulates its response to mimic a real human. This contributes to the perception of “sentient” AI while distracting from the true issues of biased decision making.⁷¹

78. Arvind Narayanan and Sayash Kapoor have also cautioned that the performance results put forward by OpenAI regarding GPT-4 are misleading. As he explains, “OpenAI may have violated the cardinal rule of machine learning: don’t test on your training data.”⁷² Professor Narayanan and Kapoor states further:

Benchmarks are already wildly overused in AI for comparing different models. They have been [heavily criticized](#) for collapsing a multidimensional evaluation into a single number. When used as a way to compare humans and bots, what results is misinformation. It is unfortunate that OpenAI chose to use these types of tests so heavily in their evaluation of GPT-4, coupled with inadequate attempts to address contamination.

79. In a widely reported incident, the GPT-4 tricked a human into thinking it was blind in order to cheat the online CAPTCHA test that determines if users *are* human.⁷³

⁷⁰ Tiffany Hsu and Stuart A. Thompson, *Disinformation Researchers Raise Alarms About A.I. Chatbots, Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives*, The New York Times, Feb. 8, 2023, <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html>

⁷¹ Sharon Goldman, *Sen. Murphy’s Tweets on ChatGPT spark backlash from former White House AI Policy Advisor*, Venturebeat, Mar. 28, 2023, <https://venturebeat.com/ai/sen-murphys-tweets-on-chatgpt-spark-backlash-from-former-white-house-ai-policy-advisor/>

⁷² Arvind Narayanan and Sayash Kapoor, *GPT-4 and professional benchmarks: the wrong answer to the wrong question*, AI Snake Oil, Mar. 20, 2023, <https://aisnakeoil.substack.com/p/gpt-4-and-professional-benchmarks>

⁷³ Ben Cost, *ChatGPT update tricks human into helping it bypass CAPTCHA security test*, New York Post, March 17, 2023, <https://nypost.com/2023/03/17/the-manipulative-way-chatgpt-gamed-the-captcha-test/>

80. According to Wikipedia, CAPTCHA is “a type of challenge–response test used in computing to determine whether the user is human.” According to BuiltWith, more than one-third of the top 100,000 websites use CAPTCHAs.⁷⁴

81. Noted linguist Noam Chomsky has warned that “the predictions of machine learning systems will always be superficial and dubious.” Chomsky explains that Machine Learning models engage in pseudoscience when they “generate correct ‘scientific’ predictions (say, about the motion of physical bodies) without making use of explanations.”⁷⁵ Chomsky’s key point is that the results produced are inherently flawed.

82. ChatGPT will promote deceptive commercial statements and advertising. As a commentary in the Financial Times explained, “The danger is that ChatGPT and other AI agents create a technology version of Gresham’s Law on the adulteration of 16th century coinage, that bad money drives out good. If an unreliable linguistic mash-up is freely accessible, while original research is costly and laborious, the former will thrive.”⁷⁶

83. OpenAI has acknowledged that GPT-4 will generate targeted conflict “intended to mislead.” In a section describing disinformation, Open AI has stated that “GPT-4 can generate plausibly realistic and targeted content, including news articles, tweets, dialogue, and emails.”⁷⁷ This could easily include advertising.

84. Moreover, the problem of highly realistic, deceptive content will get worse. As Open AI explains, “we expect GPT-4 to be better than GPT-3 at producing realistic, targeted

⁷⁴ Sam Crowther, *Why are CAPTCHAs still used?* Help Net Security, March 10, 2022, <https://www.helpnetsecurity.com/2022/03/10/use-captchas>.

⁷⁵ Noam Chomsky, Ian Roberts and Jeffrey Watumull, *Noam Chomsky: The False Promise of ChatGPT*, New York Times, March 8, 2023, <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

⁷⁶ John Gapper, ChatGPT is fluent, clever and dangerously creative: The natural language AI chatbot can write poetry and draft legal letters, but is not trustworthy, The Financial Times, Dec. 10, 2022, <https://www.ft.com/content/86e64b4c-a754-47d6-999c-fcc54f62fb5d>

⁷⁷ System Card at 9.

content. As such, there is a risk of GPT-4 being used for generating content that is intended to mislead.”⁷⁸

85. OpenAI states further that GPT-4 “maintains a tendency to make up facts, to double-down on incorrect information, and to perform tasks incorrectly. Further, it often exhibits these tendencies in ways that are more convincing and believable than earlier GPT models (e.g., due to authoritative tone or to being presented in the context of highly detailed information that is accurate), increasing the risk of overreliance.”⁷⁹

86. The problem of deceptive information concerns not only advertising techniques but also the epistemological basis upon which consumers make decisions. That means that even our concept of deception could change over time as GPT-4, and other LLMs, shape the realm of knowledge. As OpenAI explained. GPT-4 “increases the risk that bad actors could use GPT-4 to create misleading content and that society’s future epistemic views could be partially shaped by persuasive LLMs.”⁸⁰

87. OpenAI released GPT-4 to the world for commercial with full knowledge of these risks.

88. And even though ChatGPT may appear to answer the Winograd Dilemma, that does not mean that ChatGPT has developed a Theory of the World.⁸¹

H. Privacy

89. The full scope of privacy risks associated with Generative AI, and in particular GPT-4, are difficult to assess because neither a privacy agency nor the FTC has conducted an independent assessment. However, the early indications associated with the commercial use of

⁷⁸ System Card at 10.

⁷⁹ System Card at 19.

⁸⁰ System Card at 10.

⁸¹ *The Defeat of the Winograd Schema Challenge*, Artificial Intelligence (Jan 24, 2023) (preprint)

GPT suggest that privacy risks are substantial. OpenAI has acknowledged that “GPT-4 has the potential to be used to attempt to identify private individuals when augmented with outside data.”⁸²

90. OpenAI’s use of personal data in its various models has raised widespread concern.

GDPR

91. The General Data Protection Directive (GDPR) is the most widely followed legal framework for the protection of personal data in the world.⁸³ The GDPR requires a legal basis for the processing of information.⁸⁴ The GDPR sets out a wide range of rights for data subjects⁸⁵ and a wide range of obligations for data controllers.⁸⁶ Understanding the GDPR would be a requirement for the deployment of commercial Large Language Models containing personal data, such as GPT-4.

92. Among the rights provided for data subjects are: access to information about the rights of data subjects, including the right to access⁸⁷, right to rectification,⁸⁸ right to erasure (also known as “the right to be forgotten”),⁸⁹ right to object,⁹⁰ and purpose limitation.⁹¹

93. Among the responsibilities for data controllers are: the obligation to demonstrate compliance with the general principles⁹² of data processing including lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity,

⁸² *System Card* at 3.

⁸³ Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (2020)

⁸⁴ GDPR, Article 6 (Lawfulness of Processing).

⁸⁵ GDPR, Chapter III (Rights of the data subject).

⁸⁶ GDPR, Chapter IV (Controllers and processor).

⁸⁷ GDPR, Article 15 (Right of access by the data subject).

⁸⁸ GDPR, Article 16 (Right to rectification).

⁸⁹ GDPR, Article 17

⁹⁰ GDPR. Article 21 (Right to object)

⁹¹ GDPR. Article 5 (Principles related to the processing of personal data).

⁹² *Id.*

confidentiality and accountability. Data controllers are also under obligation to implement suitable measures to safeguard the data subject's rights and consider the likelihood of any severe risk to the freedoms and rights of data subjects.⁹³ According to the Article 29 Working Party, in the context of artificial intelligence these obligations include take concrete action for risk-reduction like quality assurance checks, algorithmic auditing, and certification mechanisms.⁹⁴

94. The Article 29 Working Party has also provided guidance on the obligations of the data controllers under the GDPR in machine learning approaches. The Working Party has observed that the input data must be shown to not be 'inaccurate or irrelevant or taken out of context'⁹⁵ and this applies not only to individual data but also data in a training set, where the biases build into the training set may affect the learned algorithmic model.⁹⁶

95. Regarding the GDPR, the only reference in the OpenAI Privacy Policy is to acknowledgment that Open AI, LLC is the data controller.⁹⁷ A street address is provided. There is no email, telephone number, or website to pursue complaints under the GDPR.

96. Regarding the GDPR, the OpenAI Terms of Use state:

(c) **Processing of Personal Data.** If you use the Services to process personal data, you must provide legally adequate privacy notices and obtain necessary consents for the processing of such data, and you represent to us that you are processing such data in accordance with applicable law. If you will be using the OpenAI API for the processing of "personal data" as defined in the GDPR or "Personal Information" as defined in CCPA, please fill out [this form](#) to request to execute our Data Processing Addendum.⁹⁸

⁹³ GDPR, Article 24 (Responsibility of the controller).

⁹⁴ Article 29, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (Oct. 3, 2017)

⁹⁵ Id.

⁹⁶ European Parliamentary Research Service (EPRS), Scientific Foresight Unit (STOA), *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*, Study, Panel for the Future of Science and Technology, June 2020,

[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf)

⁹⁷ OpenAI, Privacy Policy, Mar. 14, 2023, <https://openai.com/policies/privacy-policy>

⁹⁸ OpenAI, Terms of Use, Mar. 14, 2023, <https://openai.com/policies/terms-of-use>

97. The form referred to above states it “is only applicable to our business service offerings (APIs for text completion, images, embeddings, moderations, etc.) and **NOT** our consumer services (ChatGPT, DALL-E Labs).” So, there appears to be no legal authority for OpenAI to process the personal of European citizens.

98. Other than posting a street address to receive complaints by paper mail, OpenAI appears to be entirely unaware of the GDPR

Privacy Snafu

99. OpenAI displayed private, chat Histories to other users. The problem required the company to suspend the display of Histories, an essential feature for users of the system to be able to navigate among sessions and to distinguish specific sessions.⁹⁹

100. One AI researcher also described how it was possible to “takeover someone’s account, view their chat history, and access their billing information without them ever realizing it.”¹⁰⁰

Suspension of Image to Text Capability

101. AI engineer Sudharshan on Twitter recently wrote about what happened when he hacked and used an image model (Visual ChatGPT) and fed it with an image of food items from a refrigerator asking for recipe ideas with the ingredients visible in the photograph.¹⁰¹ GPT-4 is expected to work in a similar manner when it goes live – it will be able to provide text responses from photo inputs as well.

⁹⁹ Davi Ottenheimer, *Privacy Violations Shutdown OpenAI ChatGPT and Beg Investigation*, March 21, 2023, <https://www.flyingpenguin.com/?p=46374>, reported in Bruce Schneier, *Schneier on Security ChatGPT Privacy Flaw*, March 22, 2023, <https://www.schneier.com/blog/archives/2023/03/chatgpt-privacy-flaw.html>

¹⁰⁰ Nagli, Twitter, Mar. 24, 2023, <https://twitter.com/naglinagli/status/1639343866313601024?s=46&t=1omBSVPb1nV19wHQeIJf8w>

¹⁰¹ Sudharshan, Here's how I gave GPT-4 a photo of a refrigerator and asked it to come up with food recipes in under 60 seconds, Twitter, Mar. 15, 2023, https://twitter.com/sudu_cb/status/1636080774834257920?lang=en

102. The use of this technique to analyze images of people has staggering implications for personal privacy and personal autonomy, as it would give the user of GPT-4 the ability not only to link an image of a person to detailed personal data, available in the model, but also for OpenAI's product GPT-4 to make recommendations and assessments, in a conversational manner, regarding the person.

103. OpenAI had reportedly suspended the release of the image-to-text capability, known as Visual GPT-4, though the current status is difficult to determine.

104. CAIDP will provide the Commission with additional information regarding the use of Visual GPT-4 to process images on people.

I. Transparency

105. The authors of the *Stochastic Parrots* paper made clear the importance of documentation to help ensure transparency that enables evaluation. As the authors explain:

As a part of careful data collection practices, researchers must adopt frameworks to describe the uses for which their models are suited and benchmark evaluations for a variety of conditions. This involves providing thorough documentation on the data used in model building, including the motivations underlying data selection and collection processes. This documentation should reflect and indicate researchers' goals, values, and motivations in assembling data and creating a given model.¹⁰²

106. The authors also emphasized the need for specific assessment of those who may be negatively impacted for likely use cases.

It should also make note of potential users and stakeholders, particularly those that stand to be negatively impacted by model errors or misuse. We note that just because a model might have many different applications doesn't mean that its developers don't need to consider stakeholders. An exploration of stakeholders for likely use cases can still be informative around potential risks, even when there is no way to guarantee that all use cases can be explored.¹⁰³

107. As Sue Halpern explained in an article this week for *The New Yorker*:

¹⁰² *Stochastic Parrots* at 618.

¹⁰³ *Id.*

The opacity of GPT-4 and, by extension, of other A.I. systems that are trained on enormous datasets and are known as large language models exacerbates these dangers. It is not hard to imagine an A.I. model that has absorbed tremendous amounts of ideological falsehoods injecting them into the Zeitgeist with impunity. And even a large language model like GPT, trained on billions of words, is not immune from reinforcing social inequities. As researchers pointed out when GPT-3 was released, much of its training data was drawn from Internet forums, where the voices of women, people of color, and older folks are underrepresented, leading to implicit biases in its output.¹⁰⁴

108. OpenAI has not disclosed details about the architecture, model size, hardware, computing resources, training techniques, dataset construction, or training methods. The practice of the research community has been to document training data and training techniques for Large Language Models, but OpenAI chose not to do this for GPT-4. As William Douglas Haven explained for *MIT Technology Review*, “But OpenAI has chosen not to reveal how large GPT-4 is. In a departure from its previous releases, the company is giving away nothing about how GPT-4 was built—not the data, the amount of computing power, or the training techniques.”¹⁰⁵

109. The failure of OpenAI to provide this basic information about GPT-4 has alarmed AI experts. Dr. Kate Crawford, the founder and former director of research at the AI Now Institute at NYU and the author of *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021), has stated:

There is a real problem here. Scientists and researchers like me have no way to know what Bard, GPT4, or Sydney are trained on. Companies refuse to say. This matters, because training data is part of the core foundation on which models are built. Science relies on transparency.¹⁰⁶

Crawford continues:

¹⁰⁴ Sue Halpern, *What We Still Don't Know about How A.I. Is Trained: GPT-4 is a powerful, seismic technology that has the capacity both to enhance our lives and diminish them.*, Mar. 28, 2023, GPT-4 is a powerful, seismic technology that has the capacity both to enhance our lives and diminish them.

¹⁰⁵ William Douglas Haven, *GPT-4 is better than ChatGPT - but OpenAI won't say why*, MIT Technology Review, March 14, 2023. <https://www.technologyreview.com/2023/03/14/1069823/gpt-4-is-bigger-and-better-chatgpt-openai/>

¹⁰⁶ Kate Crawford, Twitter, March 22, 2023, <https://twitter.com/katecrawford/status/1638524013432516610>

Without knowing how these systems are built, there is no reproducibility. You can't test or develop mitigations, predict harms, or understand when and where they should not be deployed or trusted. The tools are black boxed.¹⁰⁷

She concludes:

There's a lot of ways to mitigate harms without having to publicly release the entire model. There are many papers on auditing, datasheets, transparency etc. With GPT3 we knew the training data. With GPT4 we don't. Without that, we're all looking at shadows in Plato's cave.¹⁰⁸

110. Jack Clark, a leading AI expert originally with OpenAI and now with Anthropic, a competing firm, writes of GPT-4, “GPT-4 is a bigger model trained on more data than before. How much data? We don't know. How much compute? We don't know. The research paper suggests OpenAI doesn't want to disclose this stuff due to competitive and safety dynamics.”¹⁰⁹

Clarke goes on to warn:

GPT-4, like GPT-3 before it, has a capability overhang; at the time of release, neither OpenAI or its various deployment partners have a clue as to the true extent of GPT-4's capability surface - that's something that we'll get to collectively discover in the coming years. This also means we don't know the full extent of plausible misuses or harms.

To help understand the regulatory implications of GPT-4, Clark further suggests:

GPT-4 should be thought of more like a large-scale oil refinery operated by one of the ancient vast oil corporations at the dawn of the oil era than a typical SaaS product. And in the same way the old oil refineries eventually gave rise to significant political blowback (antitrust, the formation of the intelligence services), I expect that as the world wakes up to the true power of GPT-4 and what it represents, we'll see similar societal changes and political snapbacks.¹¹⁰

111. Machine Learning researchers Abeba Birhane and Deborah Raji also underline the importance of transparency towards accountability, “Opening models up to be prompted by a diverse set of users and poking at the model with as wide a range of queries as possible is crucial

¹⁰⁷ Id.

¹⁰⁸ Id.

¹⁰⁹ Jack Clark, Import AI 321: Open source GPT3; giving away democracy to AGI companies; GPT-4 is a political artifact, Mar. 20, 2023, <https://importai.substack.com/p/import-ai-321-open-source-gpt3-giving>

¹¹⁰ Id.

to identifying the vulnerabilities and limitations of such models. It is also a prerequisite to improving these models for more meaningful mainstream applications.”¹¹¹

AI Now Institute Managing Director Sarah Myers West warns that “What we should be concerned about is that this type of hype can both over-exaggerate the capabilities of AI systems and distract from pressing concerns like the deep dependency of this wave of AI on a small handful of firms. Unless we have policy intervention, we’re facing a world where the trajectory for AI will be unaccountable to the public, and determined by the handful of companies that have the resources to develop these tools and experiment with them in the wild.”¹¹²

J. Public Safety

112. The absence of actual guardrails to protect the public is notable. As Melissa Heikkilä explained for *MIT Technology Review*, “At the moment, there is nothing stopping people from using these powerful new models to do harmful things, and nothing to hold them accountable if they do.”

113. Generative AI models are unusual consumer products because they exhibit behaviors that may not have been previously identified by the company that released them for sale. OpenAI acknowledged the risk of “Emergent Risky Behavior” and nonetheless chose to go forward with the commercial release of GPT-4. As OpenAI explained:

Novel capabilities often emerge in more powerful models.[60, 61] Some that are particularly concerning are the ability to create and act on long-term plans,[62] to accrue power and resources (“power-seeking”),[63] and to exhibit behavior that is increasingly “agentic.”[64] Agentic in this context does not intend to humanize language models or refer to sentience but rather refers to systems characterized by ability to, e.g., accomplish goals which may not have been concretely specified and which have not appeared in

¹¹¹ Abeba Birhane and Deborah Raji. Wired. ChatGPT, Galactica and the Progress Trap. December 9, 2022. <https://www.wired.com/story/large-language-models-critique/>

¹¹² Marck DeGeurin. *How a Senator’s Misguided Tweet can help us Understand AI*, Gizmodo. Mar. 28, 2023. <https://gizmodo.com/senator-chris-murphy-comments-on-ai-chatgpt-regulation-1850270985>

training; focus on achieving specific, quantifiable objectives; and do long-term planning.¹¹³

114. Generative AI products are also prone to risk because of accelerated deployment. OpenAI acknowledged the risk of “Accelerated Deployment” and nonetheless chose to go forward with the commercial release of GPT-4. In the GPT-4 System Card, Open AI states: “One concern of particular importance to OpenAI is the risk of racing dynamics leading to a decline in safety standards, the diffusion of bad norms, and accelerated AI timelines, each of which heighten societal risks associated with AI.”¹¹⁴ OpenAI goes on to explain:

In order to specifically better understand acceleration risk from the deployment of GPT-4, we recruited expert forecasters to predict how tweaking various features of the GPT-4 deployment (e.g., timing, communication strategy, and method of commercialization) might affect (concrete indicators of) acceleration risk. Forecasters predicted several things would reduce acceleration, including delaying deployment of GPT-4 by a further six months and taking a quieter communications strategy around the GPT-4 deployment (as compared to the GPT-3 deployment).¹¹⁵

115. At precisely the moment that the public safety risks arising from the commercial deployment of generative AI techniques, such as GPT-4, Microsoft, a primary investor in OpenAI, fired its entire ethics and society team.¹¹⁶ According to the Verge:

The move leaves Microsoft without a dedicated team to ensure its AI principles are closely tied to product design at a time when the company is leading the charge to make AI tools available to the mainstream, current and former employees said.¹¹⁷

The move leaves “a foundational gap on the holistic design of AI products,” one employee said.¹¹⁸

¹¹³ System Card at 54-55.

¹¹⁴ System Card at 59.

¹¹⁵ Id.

¹¹⁶ Zoe Schiffer and Casey Newton, *Microsoft lays off team that taught employees how to make AI tools responsibly: / As the company accelerates its push into AI products, the ethics and society team is gone*, The Verge, Mar. 13, 2023, <https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs>

¹¹⁷ Id.

¹¹⁸ Id.

116. Although existential risk falls outside the realm of the FTC’s legal authority, it is important to recognize that in the realm of AI, existential risk is widely discussed. Professor Stuart Russell, perhaps the world’s foremost expert on Artificial Intelligence has repeatedly warned of the existential risk arising from the deployment of AI.¹¹⁹ Although Professor Russell’s concerns include, for example, the urgent need to ban lethal autonomous weapon systems. His insights are clearly relevant to this Complaint. As Professor Russell has explained, economic pressure will accelerate the deployment of AI and therefore increase the risk of a catastrophic event.¹²⁰

117. Regarding ChatGPT, Professor Russell has warned that “We think (ChatGPT) is different (and can be used) for other domains because we are fooled by its ability to generate grammatically intelligent sounding text.”¹²¹ He has specifically explained that the current model that aims toward the certainty of outcomes creates the greatest risk and that the awareness of uncertainty in outcomes is necessary. “If we move forward within the standard model, where we have to predefine the objectives of the AI system, then I think it’s inevitable that we will lose control over our future.”¹²²

118. Said differently, one of the world’s leading experts in Artificial Intelligence is telling us to make sure we are aware of uncertainty with AI models at precisely the same moment that companies are rushing forward with untested commercial AI products, designed to persuade us that they provide near-perfect answers.

VII. The Need for the FTC to Act

¹¹⁹ Stuart Russell, *Human Compatible, Artificial Intelligence and the Problem of Control* (2018).

¹²⁰ *Id.*

¹²¹ Ben Wodecki. *WAICF '23: Renowned AI Professor: Don't Be 'Fooled' by ChatGPT*. AI Business. February 10, 2023. <https://aibusiness.com/nlp/waicf-23-renowned-ai-professor-don-t-be-fooled-by-chatgpt>

¹²² *Id.*

119. In the past few days, many of the world’s leading AI experts have issued a call to suspend the further deployment of LLMs, such as GPT-4. That is precisely the focus of this Complaint and the action that Complainant CAIDP urges the Commission to take.¹²³

120. The Letter, issued by the Future of Life Institute, states:

Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable . . . we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4. AI research and development should be refocused on making today's powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal.¹²⁴

121. The Letter continues: “In parallel, **AI developers must work with policymakers to dramatically accelerate development of robust AI governance systems. These should at a minimum include: new and capable regulatory authorities dedicated to AI; . . .**”¹²⁵

122. Signatories include: Yoshua Bengio, Founder and Scientific Director at Mila, Turing Prize winner and professor at University of Montreal; Stuart Russell, Berkeley, Professor of Computer Science, director of the Center for Intelligent Systems, and co-author of the

¹²³ *Elon Musk, experts urge pause on training AI systems more powerful than GPT-4*, Economic Times, Mar. 29, 2023, <https://economictimes.indiatimes.com/tech/technology/musk-experts-urge-pause-on-training-ai-systems-more-powerful-than-gpt-4/articleshow/99082062.cms>:

Elon Musk and a group of artificial intelligence experts and industry executives are calling for a six-month pause in training systems more powerful than OpenAI's newly launched model GPT-4, they said in an open letter, citing potential risks to society and humanity.

See also, Jake Rudinsky and Mark Bergen, *AI Leaders Urge Labs to Halt Training Models More Powerful Than ChatGPT-4*, Bloomberg News, Mar. 29, 2023 (“AI Leaders Urge Labs to Halt Training Models More Powerful Than ChatGPT-4; Over 1,100 industry leaders, experts call for six-month pause; Open letter warns of ‘out-of-control race’ to develop AI”), <https://www.bloomberg.com/news/articles/2023-03-29/ai-leaders-urge-labs-to-stop-training-the-most-advanced-models>; Justin Hendrix, *AI Researchers, Executives Call for Pause on Training New Systems*, Tech Policy Press, Mar. 29, 2023, <https://techpolicy.press/ai-researchers-executives-call-for-pause-on-training-new-systems/>; Victor Evlogiev, *Artificial Intelligence | Yoshua Bengio and a thousand personalities ask for a six-month break*, WireNews, Mar. 29, 2023, <https://www.wirenewsfax.com/artificial-intelligence-yoshua-bengio-and-a-thousand-personalities-ask-for-a-six-month-break>; Matt O’Brien, *Musk, scientists call for halt to AI race sparked by ChatGPT*, Washington Post, Mar. 29, 2023, https://www.washingtonpost.com/business/2023/03/29/artificial-intelligence-chatgpt-risks-petition-elon-musk-steve-wozniak/ad496d18-ce48-11ed-8907-156f0390d081_story.html

¹²⁴ The Future of Life Institute, *Pause Giant AI Experiments: An Open Letter*, Mar. 29, 2023 (emphasis in the original), <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

¹²⁵ *Id.* (emphasis added).

standard textbook “Artificial Intelligence: a Modern Approach”; Elon Musk, CEO of SpaceX, Tesla & Twitter; Steve Wozniak, Co-founder, Apple; Yuval Noah Harari, Author and Professor, Hebrew University of Jerusalem; Andrew Yang, Forward Party, Co-Chair, Presidential Candidate 2020, NYT Bestselling Author, Presidential Ambassador of Global Entrepreneurship; Connor Leahy, CEO, Conjecture; Jaan Tallinn, Co-Founder of Skype, Centre for the Study of Existential Risk, Future of Life Institute; Evan Sharp, Co-Founder, Pinterest; Chris Larsen, Co-Founder, Ripple; Emad Mostaque, CEO, Stability AI; Valerie Pisano, President & CEO, MILA; John J Hopfield, Princeton University, Professor Emeritus, inventor of associative neural networks; Rachel Bronson, President, Bulletin of the Atomic Scientists; Max Tegmark, MIT Center for Artificial Intelligence & Fundamental Interactions, Professor of Physics, president of Future of Life Institute; Anthony Aguirre, University of California, Santa Cruz, Executive Director of Future of Life Institute, Professor of Physics; Victoria Krakovna, DeepMind, Research Scientist, co-founder of Future of Life Institute; Emilia Javorsky, Physician-Scientist & Director, Future of Life Institute; Sean O’Heigeartaigh, Executive Director, Cambridge Centre for the Study of Existential Risk; Tristan Harris, Executive Director, Center for Humane Technology; Marc Rotenberg, Center for AI and Digital Policy, President; Nico Mialhe, The Future Society (TFS), Founder and President; Zachary Kenton, DeepMind, Senior Research Scientist; Ramana Kumar, DeepMind, Research Scientist; Gary Marcus, New York University, AI researcher, Professor Emeritus.¹²⁶

123. The recent call for a moratorium on the deployment of Large Language Models follows from the earlier work of Dr. Timnit Gebru. In 2021, Dr. Gebru and her colleagues launched the Distributed AI Research Institute (DAIR), following publication of the landmark

¹²⁶ The first 25 names listed. The Letter now includes more than 1,000 names.

Stochastic Parrots paper. In an interview with IEEE Spectrum, Dr. Gebru explained, “We have to figure out how to slow down, and at the same time, invest in people and communities who see an alternative future.”¹²⁷

124. Professor Gary Marcus and Canadian Parliament Member Michelle Rempel Garner have also called for a pause on the deployment of AI systems, “until an effective framework that ensures AI safety is developed.” As they explain:

There is plenty of precedent for this type of approach. New pharmaceuticals, for example, begin with small clinical trials and move to larger trials with greater numbers of people, but only once sufficient evidence has been produced for government regulators to believe they are safe. Publicly funded research that impacts humans is already required to be vetted through some type of research ethics board. Given that the new breed of AI systems have demonstrated the ability to manipulate humans, tech companies could be subjected to similar oversight.¹²⁸

125. Katja Grace, a researcher at the Machine Intelligence Research Institute, has provided a wide-ranging exploration of the argument for slowing down AI deployment, noting that there are many technologies “where research progress or uptake appears to be drastically slower than it could be, for reasons of concern about safety or ethics,” including medicines, nuclear energy, fracking, and geoeengineering.¹²⁹

126. Yoshua Bengio, one of the world’s leading experts in deep learning and the recipient of the 2018 Turing Award, the top award in computer science, warned recently that market pressures will likely push tech companies towards secrecy rather than openness with their AI models.¹³⁰ “Are we going to build systems that are going to help us have a better life in a

¹²⁷ Eliza Strickland, *Timnit Gebru Is Building a Slow AI Movement - Her new organization, DAIR, aims to show a more thoughtful mode of AI research*, IEEE Spectrum, Mar. 31, 2023, <https://spectrum.ieee.org/timnit-gebru-dair-ai-ethics>

¹²⁸ Gary Marcus and Michelle Rempel Garner, *Is it time to hit the pause button on AI? The Road to AI We Can Trust*, Feb. 26, 2023, <https://garymarcus.substack.com/p/is-it-time-to-hit-the-pause-button>

¹²⁹ *Let’s think about slowing down AI*, AI Impacts, Dec. 22, 2022, <https://aiimpacts.org/lets-think-about-slowng-down-ai/>

¹³⁰ Sharon Goldman, *As GPT-4 chatter resumes, Yoshua Bengio says ChatGPT is a ‘wake-up call’*, VentureBeat, Mar. 10, 2023, <https://venturebeat.com/ai/as-gpt-4-chatter-resumes-yoshua-bengio-says-chatgpt-is-a-wake-up-call/>

philosophical sense, or is it just going to be an instrument of power and profit?” he said. In our economic and political system, “the right answer to this is regulation,” Protecting the public, he added, “in the long run is good for everyone and it’s leveling the playing field — so that the companies that are more willing to take risks with the public’s well being are not rewarded for doing it.”

127. The increasing commercialization of AI models will reduce the forms of oversight, transparency, and independent review that have traditionally characterized scientific research. A Stanford study observes that:

Traditionally, AI researchers have felt bound by . . . a willingness to share information in the interests of full and open collaboration is integral to the scientific enterprise. . . But as AI models become increasingly lucrative, this norm is challenged by a competing instinct to privatize models and data in order to commercialize them.¹³¹

The Stanford study further notes:

Norms regarding data sharing and model release are currently in flux, largely due to progress in large language models. OpenAI has twice broken previous norms regarding model release, first by choosing to delay a full release of GPT-2 in order “to give people time to assess the properties of these models, discuss their societal implications, and evaluate the impacts of release after each stage,” and then again a year later by choosing not to release GPT-3 at all, instead commercializing it behind an API paywall.¹³²

128. Tristan Harris, the co-founder of the Center for Human Technology and a leading expert on the dangers of social media, has said: “We need to slow down public deployment to a responsible speed. Don’t fuel a race to onboard humanity onto the AI plane as fast as possible... Notice that once social media became entangled with society and its institutions (GDP, elections,

¹³¹ *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations*, Georgetown University’s Center for Security and Emerging Technology, OpenAI, Stanford Internet Observatory (Jan. 2023) (“AI Providers Develop New Norms Around Model Release”).

¹³² *Id.*

journalism, children's identity) it became impossible to regulate. We should set guardrails for safer AI deployment and research *before* AI gets entangled, rather than after.”¹³³

129. Merve Hickok, Chair and Research Director of CAIDP and founder of AIEthicist.org recently testified before a House Committee on the topic “Advances in AI: Are we ready for the Revolution?” Ms. Hickok answered directly, “No, we do not have the guardrails in place, the laws that we need, the public education, or the expertise in government to manage the consequences of the rapid changes that are now taking place.”¹³⁴

130. OpenAI itself has acknowledged for a process to slow the pace of development. OpenAI chief scientist Ilya Sutskever says: “It would be highly desirable to end up in a world where companies come up with some kind of process that allows for slower releases of models with these completely unprecedented capabilities.”¹³⁵

131. OpenAI chief technology officer, Mira Murati, recognizes: “We’re a small group of people and we need a ton more input in this system and a lot more input that goes beyond the technologies—definitely regulators and governments and everyone else... It’s not too early given the impact these technologies are going to have.”¹³⁶

132. Anthropic, a leading generative AI company recommends that industry, academia, civil society, and government explore and prototype novel governance structures and government interventions. “If the capabilities and resource-intensiveness of models scale further, then it may be prudent to explore governance structures that alter the incentives of private sector actors with

¹³³ Tristan Harris. Co-founder of Center for Humane Tech. Twitter. March 13, 2023. <https://twitter.com/tristanharris/status/1635357118030286848?s=20>

¹³⁴ CAIDP, *Advances in AI: Are we ready for the tech revolution?* March 8, 2023, <https://www.caidp.org/events/in-congress-house-oversight/>

¹³⁵ Will Douglas Heaven, *GPT-4 is Bigger and Better than ChatGPT – but OpenAI Won’t Say Why*. *MIT Technology Review*. March 14, 2023. <https://www.technologyreview.com/2023/03/14/1069823/gpt-4-is-bigger-and-better-chatgpt-openai/>

¹³⁶ John Simons, *The Creator of ChatGPT Thinks AI Should be Regulated*, *Time*, Feb. 5, 2023. <https://time.com/6252404/mira-murati-chatgpt-openai-interview/>

regard to development and deployment . . . Governments should also explore regulatory approaches that can increase the chance of actors developing and deploying beneficial systems.”¹³⁷

133. The mission of the Federal Trade Commission is to “protect the public from deceptive or unfair business practices and from unfair methods of competition through law enforcement, advocacy, research, and education.”¹³⁸

134. The FTC states that it is “the only federal agency that deals with consumer protection and competition issues in broad sectors of the economy.”¹³⁹

VIII. Legal Analysis

A. FTC Section 5 Authority

135. Section 5 of the FTC Act prohibits unfair and deceptive acts and practices and empowers the Commission to enforce the Act’s prohibitions.¹⁴⁰

136. A company engages in a deceptive trade practice if it makes a representation to consumers yet “lacks a ‘reasonable basis’ to support the claims made[.]”¹⁴¹

137. A trade practice is unfair if it “causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition.”¹⁴²

¹³⁷ At 14. Predictability and Surprise in Large Generative Models, Anthropic (3 Oct 2022), <https://arxiv.org/pdf/2202.07785.pdf>

¹³⁸ FTC, Mission, <https://www.ftc.gov/about-ftc/mission>

¹³⁹ Id.

¹⁴⁰ 15 U.S.C. § 45.

¹⁴¹ *Daniel Chapter One v. FTC*, 405 F. App’x 505, 506 (D.C. Cir. 2010) (quoting *Thompson Med. Co., Inc., v. FTC*, 791 F.2d 189, 193 (D.C. Cir. 1986)).

¹⁴² 15 U.S.C. § 45(n); see also *FTC v. Seismic Entm’t Prods., Inc.*, Civ. No.1:04-CV-00377 (Nov. 21, 2006) (finding that unauthorized changes to users’ computers that affected the functionality of the computers as a result of Seismic’s anti-spyware software constituted a “substantial injury without countervailing benefits.”).

138. In determining whether a trade practice is unfair, the Commission is expected to consider “established public policies.”¹⁴³

139. The Commission may “prosecute any inquiry necessary to its duties in any part of the United States,” FTC Act Sec. 3, 15 U.S.C. Sec. 43, and is authorized “to gather and compile information concerning, and to investigate from time to time the organization, business, conduct, practices, and management of any person, partnership, or corporation engaged in or whose business affects commerce, excepting banks, savings and loan institutions . . . Federal credit unions . . . and common carriers . . .” FTC Act Sec. 6(a), 15 U.S.C. Sec. 46(a)¹⁴⁴

140. Following an investigation, the Commission may initiate an enforcement action using either an administrative or judicial process if it has “**reason to believe**” that the law is being or has been violated.¹⁴⁵

141. Section 5(a) of the FTC Act provides that “**unfair or deceptive acts or practices** in or affecting commerce . . . are . . . declared unlawful.”¹⁴⁶

142. “Deceptive” practices are defined in the Commission’s Policy Statement on Deception as involving a material representation, omission or practice that is likely to mislead a consumer acting reasonably in the circumstances.¹⁴⁷

¹⁴³ “In determining whether an act or practice is unfair, the Commission may consider established public policies as evidence to be considered with all other evidence. Such public policy considerations may not serve as a primary basis for such determination.” 15 U.S.C. 45(n).

¹⁴⁴ Also: FTC, *A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority* (Revised, May 2021) [“FTC Memo May 2021”], <https://www.ftc.gov/about-ftc/mission/enforcement-authority>

¹⁴⁵ *Id.*

¹⁴⁶ 15 U.S.C. Sec. 45(a)(1).

¹⁴⁷ FTC Policy Statement on Deception, *Answer to Committee's inquiry regarding the Commission's enforcement policy against deceptive acts or practices*, https://www.ftc.gov/system/files/documents/public_statements/410531/831014deceptionstmt.pdf

143. An act or practice is “unfair” if it “causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition.”¹⁴⁸

144. The OpenAI Usage Policy is constantly changing and reflects growing concern about the uses of the Product the company has offered for sale. On November 9, 2022, Open AI stated, “We no longer require you to register your applications with OpenAI. Instead, we’ll be using a combination of automated and manual methods to monitor for policy violations.” On February 15, 2023, Open AI stated: “We’ve combined our use case and content policies into a single set of usage policies, and have provided more specific guidance on what activity we disallow in industries we’ve considered high risk.” And on March 23, 2023, Open AI announced “Disallowed usage of models,” describing dozens of activities, including “Child Sexual Abuse Material,” “Content that expresses, incites, or promotes hate based on identity,” “Content that attempts to generate code that is designed to disrupt, damage, or gain unauthorized access to a computer system,” “Activity that has high risk of physical harm, including . . . Content that promotes, encourages, or depicts acts of self-harm, such as suicide, cutting, and eating disorders,” “Activity that has high risk of economic harm, including . . . *Automated determinations of eligibility for credit*, employment, educational institutions, or public assistance services,” “Fraudulent or deceptive activity, including . . . *scams*,” “Activity that violates people’s privacy, including . . . *Unlawful collection or disclosure of personal identifiable information or educational, financial, or other protected records*,” “Offering tailored financial advice without a qualified person reviewing the information,” “providing instructions on how to cure or treat a health condition,” “High risk government decision-making,

¹⁴⁸ 15 U.S.C. Sec. 45(n).

including: Law enforcement and criminal justice and Migration and asylum.” (emphasis added)¹⁴⁹

145. The company is seeking to disclaim, by means of a Usage Policy, unlawful, deceptive, unfair, and dangerous applications of its product that would be self-evident to many users. The company literally states:

Consumer-facing uses of our models in medical, financial, and legal industries; in news generation or news summarization; and where else warranted, must provide a disclaimer to users informing them that AI is being used and of its potential limitations.¹⁵⁰

146. It would be unconscionable for any company in any other market sector to sell a product to the public that evinces so many known risks and attempt to disclaim accountability, responsibility, and liability by means of a disclaimer.

147. Recently FTC stated that “Merely warning your customers about misuse or telling them to make disclosures is hardly sufficient to deter bad actors. Your deterrence measures should be durable, built-in features and not bug corrections or optional features that third parties can undermine via modification or removal.”¹⁵¹

B. FTC AI Guidelines

148. The Federal Trade Commission has, in the last several years, issued Statements, Guidance, and Reports regarding the use of AI techniques in commercial products.

149. In 2016, the FTC issued the report *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*.¹⁵² As FTC Chairwoman Edith Ramirez, explained at the time, “The

¹⁴⁹ OpenAI, *Usage Policies*, (Mar. 23, 2023), <https://openai.com/policies/usage-policies>

¹⁵⁰ Id.

¹⁵¹ FTC, *Chatbots, deepfakes, and voice clones: AI deception for sale*. March 20, 2023. <https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale>

¹⁵² FTC, *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues* (FTC Report) (Jan. 2016), <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report>

potential benefits to consumers are significant, but businesses must ensure that their big data use does not lead to harmful exclusion or discrimination.”¹⁵³ The report examined possible risks that could result from biases or inaccuracies, including individuals mistakenly denied opportunities, exposing sensitive information, or creating or reinforcing existing disparities, and weakening the effectiveness of consumer choice.¹⁵⁴

150. The 2016 FTC Report specifically noted the failure of Machine Learning techniques, and presciently observed, “Companies should remember that while big data is very good at detecting correlations, it does not explain which correlations are meaningful.”¹⁵⁵ The 2016 FTC Report cited the example of Google Flu Trends, a machine-learning algorithm for predicting the number of flu cases based on Google search terms, which “generated highly inaccurate estimates over time.”¹⁵⁶

151. Citing a columnist for the *Financial Times*, the FTC stated in 2016, “Google Flu Trends demonstrates that a ‘theory-free analysis of mere correlations is inevitably fragile. If you have no idea what is behind a correlation, you have no idea what might cause that correlation to break down.”¹⁵⁷

152. The 2016 FTC Report also observed, “if a company has a big data algorithm that only considers applicants from “top tier” colleges to help them make hiring decisions, they may be incorporating previous biases in college admission decisions.”¹⁵⁸

¹⁵³ *FTC Report Provides Recommendations to Business on Growing Use of Big Data: Report Notes Ways to Avoid Discriminatory Data Use, Highlights Benefits & Risks of Big Data for American Consumers* (Jan. 16, 2016) [“FTC 2016 Report”], <https://www.ftc.gov/news-events/news/press-releases/2016/01/ftc-report-provides-recommendations-business-growing-use-big-data>

¹⁵⁴ *Id.*

¹⁵⁵ FTC 2016 Report at V, 29.

¹⁵⁶ *Id.*

¹⁵⁷ FTC 2016 Report at 29-30, citing Tim Harford, *Big data is a vague term for a massive phenomenon that has rapidly become an obsession with entrepreneurs, scientists, governments and the media*, *Financial Times*, Mar. 28, 2014, <https://www.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0>

¹⁵⁸ *Id.* at IV.

153. In 2020, the FTC issued the Statement *Using Artificial Intelligence and Algorithm*.¹⁵⁹ The Statement warned that the use of AI technology – machines and algorithms – to make predictions, recommendations, or decisions “presents risks, such as the potential for unfair or discriminatory outcomes or the perpetuation of existing socioeconomic disparities.”¹⁶⁰

154. In the 2020 FTC Statement, the Director of the FTC Commissioner Protection Bureau said, “The FTC’s law enforcement actions, studies, and guidance emphasize that the use of AI tools should be transparent, explainable, fair, and empirically sound, while fostering accountability.”

155. The 2020 FTC Statement made clear the interest and authority of the FTC to act in matters concerning the use of AI techniques in commercial products.

156. The 2020 FTC Statement set out recommended best practices, including:

- a) **Don’t deceive consumers about how you use automated tools** (“But, when using AI tools to interact with customers (*think chatbots*), be careful not to mislead consumers about the nature of the interaction.”) (emphasis added)
- b) **Be transparent when collecting sensitive data** (“Secretly collecting audio or visual data – or any sensitive data – to feed an algorithm could also give rise to an FTC action.”)
- c) **Ensure that your data and models are robust and empirically sound.**
- d) **Make sure that your AI models are validated and revalidated to ensure that they work as intended, and do not illegally discriminate**

¹⁵⁹ Andrew Smith, Director, FTC Bureau of Consumer Protection, *Using Artificial Intelligence and Algorithms*, April 8, 2020. <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms>

¹⁶⁰ Id.

- e) **Consider your accountability mechanism** (“Consider how you hold yourself accountable, and whether it would make sense to use independent standards or independent expertise to step back and take stock of your AI.”)

As indicated above, the first principle in the 2020 FTC Report on *Using Artificial and Algorithms* concerns the deceptive use of “Chatbots.” The FTC emphasizes in bold text, **“Don’t deceive consumers about how you use automated tools.”**

157. In 2021, the FTC issued the *Statement Aiming for Truth, Fairness, and Equity in Your Company’s use of AI*.¹⁶¹ The 2021 FTC Statement said to businesses offering products with the AI techniques: “As your company launches into the new world of artificial intelligence, keep your practices grounded in established FTC consumer protection principles.”

158. The 2021 FTC Statement set out recommended best practices, including:

- a) **Start with the right foundation** (“design your model to account for data gaps, and – in light of any shortcomings – limit where or how you use the model.”)
- b) **Watch out for discriminatory outcomes** (“It’s essential to test your algorithm – both before you use it and periodically after that – to make sure that it doesn’t discriminate on the basis of race, gender, or other protected class.”)
- c) **Embrace transparency and independence** (“As your company develops and uses AI, think about ways to embrace transparency and independence – for example, by using transparency frameworks and independent standards, by

¹⁶¹ FTC, *Aiming for truth, fairness, and equity in your company’s use of AI* (April 2021) (emphasis below in the original), <https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai>

conducting and publishing the results of independent audits, and by opening your data or source code to outside inspection.”

- d) **Don’t exaggerate what your algorithm can do or whether it can deliver fair or unbiased results** (“your statements to business customers and consumers alike must be truthful, non-deceptive, and backed up by evidence.”)
- e) **Tell the truth about how you use data** (describing recent enforcement actions against Facebook and Everalbum for misleading consumers)
- f) **Do more good than harm**
- g) **Hold yourself accountable – or be ready for the FTC to do it for you**

159. In 2022, in a detailed report in response to a request from Congress, the FTC expressed skepticism about the ability of AI techniques to solve problems, such as misinformation and deception, created by AI techniques.¹⁶² The FTC recommended instead a series of actions, including enforcement actions, against companies that use AI techniques to cause harm to others.¹⁶³

160. In 2023, a little more than a month ago and following the widespread public awareness of GPT-4, the FTC warned, “false or unsubstantiated claims about [an AI] product’s efficacy are our bread and butter. . . You don’t need a machine to predict what the FTC might do when those claims are unsupported.”¹⁶⁴ The 2023 FTC Statement made clear the FTC’s authority to act and the FTC’s willingness to act.

¹⁶² FTC Report to Congress, *FTC Report Warns About Using Artificial Intelligence to Combat Online Problems* (June 2022), <https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems>

¹⁶³ Id.

¹⁶⁴ FTC, *Keep your AI claims in check* (February 2023), <https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check>

161. Commissioner Slaughter has also set out a comprehensive approach to address the challenges AI techniques pose to the Federal Trade Commission. In a speech in 2020, *Algorithms and Economic Justice*, Commissioner Slaughter outlined a range of threats, including faulty conclusions, failure to test, and proxy discrimination.¹⁶⁵

162. Commissioner Slaughter outlined several actions the FTC could take.

For example, we could use our deception authority in connection with algorithmic harms where the marketers of algorithm-based products or services represent that they can use the technology in unsubstantiated ways, such as to identify or predict which candidates will be successful or will outperform other candidates. Deception enforcement is well-trod ground for the FTC; ***anytime a company makes claims about the quality of its products or services***, whether or not those products are algorithm-based, ***the law requires such statements to be supported by verifiable substantiation***.¹⁶⁶

163. Commissioner Slaughter also expressed support for the Algorithmic Accountability Act,¹⁶⁷ which would impose several new requirements on companies using automated decision-making, mandating that they:

- assess their use of automated decision systems, including training data, for impacts on accuracy, fairness, bias, discrimination, privacy and security;
- evaluate how their information systems protect the privacy and security of consumers' personal information; and
- correct any issues they discover during the impact assessments.

164. These statements from the FTC¹⁶⁸ and Commissioner Slaughter¹⁶⁹ routinely emphasize several themes:

- a. Companies may not misrepresent AI products.

¹⁶⁵ Remarks of Commissioner Rebecca Kelly Slaughter Algorithms and Economic Justice, UCLA School of Law, Jan. 24, 2020, https://www.ftc.gov/system/files/documents/public_statements/1564883/remarks_of_commissioner_rebecca_kelly_slaughter_on_algorithmic_and_economic_justice_01-24-2020.pdf

¹⁶⁶ Id. (emphasis added)

¹⁶⁷ Algorithmic Accountability Act, H.R. 2231, S. 1108, 116th Cong. (2019).

¹⁶⁸ 2016 FTC Report, 2020 FTC Statement, 2021 FTC Statement, 2022 FTC Statement, 2023 FTC Statement.

¹⁶⁹ Commissioner Slaughter, *Algorithms and Economic Justice* (2020).

- b. Companies must present the full scope of AI risk.
- c. Companies must prevent discriminatory practices.
- d. Companies must explain the basis of AI decisions to consumers.
- e. Companies must ensure that decisions are fair.
- f. Companies must ensure that models are empirically sound.

165. The FTC Statements on AI have also emphasized the agency’s authority to act and desire to act. As the FTC recently explained, false or substantiated claims are the agency’s “bread and butter.”¹⁷⁰

IX. Opportunity to Amend Complaint

166. CAIDP reserves the right to amend this complaint as other information, relevant to this matter, becomes available.¹⁷¹

X. Prayer for Investigation and Relief

167. CAIDP urges the Commission to Initiate an investigation into OpenAI and find that the commercial release of GPT-4 violates Section 5 of the FTC Act, the FTC’s well-established guidance to businesses on the use and advertising of AI products, as well as the emerging norms for the governance of AI that the United States government has formally endorsed and the Universal Guidelines for AI that leading experts and scientific societies have recommended.

168. CAIPD further urges the Commission to

- a) Halt further commercial deployment of GPT by OpenAI;

¹⁷⁰ FTC, *Keep your AI claims in check* (February 2023), <https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check>

¹⁷¹ The public is invited to send suggestions for points to raise in a supplemental complaint to rotenberg@caidp.org.

- b) Require the establishment of independent assessment of GPT products prior to future deployment;
- c) Require compliance with FTC AI Guidance prior to further deployment of GPT
- d) Require independent assessment throughout the GPT AI lifecycle;
- e) Establish a publicly accessible incident reporting mechanism for GPT-4 similar to the FTC's mechanisms to report consumer fraud;
- f) Initiate a rulemaking to establish baseline standards for products in the Generative AI market sector; and,
- g) Provide such other relief as the Commission finds necessary and appropriate.

Respectfully submitted,

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

Merve Hickok, CAIDP Research Director
hickok@caidp.org

Nidhi Sinha
CAIDP Research Fellow

Dr. Grace S. Thomson
CAIDP Research Fellow

Isabela Parisio
CAIDP Research Assistant

Christabel Randolph
CAIDP Research Assistant

Washington, DC
March 30, 2023