**GLANCY PRONGAY & MURRAY LLP**
KEVIN F. RUF (#136901)
1925 Century Park East, Suite 2100
Los Angeles, CA 90067
Telephone: (310) 201-9150
Facsimile: (310) 201-9160
Email: info@glancylaw.com

*Counsel for Plaintiff and the Proposed Classes*

**UNITED STATES DISTRICT COURT**
**NORTHERN DISTRICT OF CALIFORNIA**

A.S., individually and on behalf of all others similarly situated,

Plaintiff,

vs.

OPENAI LP, OPENAI INCORPORATED, OPENAI GP, LLC, OPENAI STARTUP FUND I, LP, OPENAI STARTUP FUND GP I, LLC, OPENAI STARTUP FUND MANAGEMENT LLC, MICROSOFT CORPORATION and DOES 1 through 20, inclusive,

Defendants.

Case No.:

**CLASS ACTION COMPLAINT**

1. VIOLATION OF ELECTRONIC COMMUNICATIONS PRIVACY ACT, 18 U.S.C. §§ 2510, *et seq.*
2. VIOLATION OF THE COMPUTER FRAUD AND ABUSE ACT, 18 U.S.C. § 1030
3. VIOLATION OF THE CALIFORNIA INVASION OF PRIVACY ACT ("CIPA"), CAL. PENAL CODE § 631
4. VIOLATION OF CALIFORNIA UNFAIR COMPETITION LAW, BUSINESS AND PROFESSIONS CODE §§ 17200, *et seq.*
5. NEGLIGENCE
6. INVASION OF PRIVACY
7. INTRUSION UPON SECLUSION
8. LARCENY/RECEIPT OF STOLEN PROPERTY
9. CONVERSION
10. UNJUST ENRICHMENT

**DEMAND FOR JURY TRIAL**

Plaintiff A.S. (hereinafter "Plaintiff"),[1] individually and on behalf of all others similarly situated, brings this action against Defendants OpenAI, OpenAI Incorporated, OpenAI GP LLC, OpenAI Startup Fund I, LP, OpenAI Startup Fund GP I, LLC, and Microsoft Corporation (collectively, "Defendants"). Plaintiff's allegations are based upon personal knowledge as to herself and her own acts, and upon information and belief as to all other matters based on the investigation conducted by and through Plaintiff's attorneys.

## INTRODUCTION

1. On October 19, 2016, University of Cambridge Professor of Theoretical Physics Stephen Hawking predicted, "Success in creating AI could be the biggest event in the history of our civilization. But it could also be the last, unless we learn how to avoid the risks."[2] Professor Hawking described a future in which humanity would choose to either harness the huge potential benefits or succumb to the dangers of AI, emphasizing "the rise of powerful AI will be either the best or the worst thing ever to happen to humanity."

2. The future Professor Hawking predicted has arrived in just seven short years. Using stolen and misappropriated personal information at scale, Defendants have created powerful and wildly profitable AI and released it into the world without regard for the risks. In so doing, Defendants have created an AI arms race in which Defendants and other Big Tech companies are onboarding society into a plane that over half of the surveyed AI experts believe has at least a 10% chance of crashing and killing everyone on board.[3] Humanity is now faced with the two choices:

---

[1] Plaintiff respectfully requests that the Court permit them to keep their identity private as Plaintiff aims to avoid intrusive scrutiny as well as any potentially dangerous backlash. Indeed, plaintiffs in other lawsuits against the same defendant entities have received many troubling and violent threats, including death threats, marking a severe infringement of personal safety. Accordingly, opting for privacy is a critical measure to avoid unwarranted negative attention as well as potential harm. Plaintiff will file a motion to proceed pseudonymously, if required. *See* Victoria Hudgins, *GitHub and Openai Plaintiffs Seek Anonymity amid Slurs and Death Threats*, GLOB. DATA REV. (Mar. 15, 2023), globaldatareview.com/article/github-and-openai-plaintiffs-seek-anonymity-amid-slurs-and-death-threats.

[2] Cambridge University, *The Best or Worst Thing to Happen to Humanity*, YOUTUBE (Oct. 19, 2016), https://www.youtube.com/watch?v=_5XvDCjrdXs&t=1s.

[3] Yuval Harari et al., *You Can Have the Blue Pill or the Red Pill, and We're Out of Blue Pills*, THE N.Y. TIMES (Mar. 24, 2023), https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html ("[O]ver 700 top academics and researchers behind the leading artificial intelligence companies were asked in a survey about future A.I. risk. Half of those surveyed stated that there was a 10 percent or greater chance of human extinction (or similarly permanent and severe disempowerment) from future A.I. systems.").

One leads to sustainability, security, and prosperity; the other leads to civilizational collapse.

3. This class action lawsuit arises from Defendants' unlawful and harmful conduct in developing, marketing, and operating their AI products, including ChatGPT-3.5, ChatGPT-4.0,[4] Dall-E, and Vall-E (the "Products"), which use stolen private information, including personally identifiable information, from hundreds of millions of internet users, including children of all ages, without their informed consent or knowledge. Furthermore, Defendants continue to unlawfully collect and feed additional personal data from millions of unsuspecting consumers worldwide, far in excess of any reasonably authorized use, in order to continue developing and training the Products.

4. Defendants' disregard for privacy laws is matched only by their disregard for the potentially catastrophic risk to humanity. Emblematic of both the ultimate risk—and Defendants' open disregard—is this statement from Defendant OpenAI's CEO Sam Altman: "AI will probably most likely lead to the end of the world, but in the meantime, there'll be great companies."[5]

5. Defendants' Products, and the technology on which they are built, have the potential to do much good in the world, like aiding life-saving scientific research and ushering in discoveries that can improve the lives of everyday Americans. With that potential in mind, Defendant OpenAI was originally founded as a nonprofit research organization with a single mission: to create and ensure artificial intelligence would be used for the benefit of humanity. But in 2019, OpenAI abruptly restructured itself, developing a for-profit business that would pursue commercial opportunities of staggering scale.

6. As a result of the restructuring, OpenAI abandoned its original goals and principles, electing instead to pursue profit at the expense of privacy, security, and ethics. It doubled down on a strategy to secretly harvest massive amounts of personal data from the internet, including private

[4] ChatGPT is referred to herein as inclusive of both ChatGPT-3.5, ChatGPT-4, and any other versions of ChatGPT. The term "ChatGPT Plug-In" encompasses GPT-3.5, GPT-4, and any additional extensions that have been incorporated into Microsoft's and third-party platforms, websites, applications, programs, or systems.

[5] Matt Weinberger, *Head of Silicon Valley's Most Important Startup Farm Says We're in A 'Mega Bubble' That Won't Last*, BUS. INSIDER (June 4, 2015), https://www.businessinsider.com/sam-altman-y-combinator-talks-mega-bubble-nuclear-power-and-more-2015-6?r=US; David Wallace-Wells, *A.I. Is Being Built by People Who Think It Might Destroy Us*, THE N.Y. TIMES (Mar. 27, 2023), https://www.nytimes.com/2023/03/27/opinion/ai-chatgpt-chatbots.html.

information and private conversations, medical data, information about children—essentially every piece of data exchanged on the internet it could take—without notice to the owners or users of such data, much less with anyone's permission.

7. Without this unprecedented theft of private and copyrighted information belonging to real people, communicated to unique communities, for specific purposes, targeting specific audiences, the Products would not be the multi-billion-dollar business they are today. OpenAI used the stolen data to train and develop the Products utilizing large language models (LLMs) and deep language algorithms to analyze and generate human-like language that can be used for a wide range of applications, including chatbots, language translation, text generation, and more. Defendants' Products' sophisticated natural language processing capabilities allow them to, among other things, carry on human-like conversations with users, answer questions, provide information, generate next text on demand, create art, and connect emotionally with people, all like a "real" human.

8. Once focused on stolen data, Defendants saw the immediate profit potential and rushed the Products to market without implementing proper safeguards or controls to ensure that they would not produce or support harmful or malicious content and conduct that could further violate the law, infringe rights, and endanger lives. Without these safeguards, the Products have already demonstrated their ability to harm humans, in real ways.

9. A nontrivial number of experts claim the risks to humanity presented by the Products outweigh even those of the Manhattan Project's development of nuclear weapons. Historically, the unchecked release of new technologies without proper safeguards and regulations has caused chaos.[6] Now again, we face imminent and unreasonable risks of the very fabric of our society

[6] Bill Kovarik, *A Century of Tragedy: How the Car and Gas Industry Knew About The Health Risks of Leaded Fuel But Sold it For 100 Years Anyway*, THE CONVERSATION (Dec. 8, 2021), https://theconversation.com/a-century-of-tragedy-how-the-car-and-gas-industry-knew-about-the-health-risks-of-leaded-fuel-but-sold-it-for-100-years-anyway-173395 (1920s invention of leaded gasoline, initially thought of as a technological breakthrough, resulted in serious health and environmental consequences, such as lead poisoning and soil contamination); James H. Kim & Anthony R. Scialli, *Thalidomide: The Tragedy of Birth Defects and the Effective Treatment of Disease*, 122 TOXICOLOGICAL SCI. 1, 1 (2011) (Development of thalidomide in the 1950s and 60s, thought to be the miraculous solution to nausea, led to widespread birth defects in babies whose mothers had taken the drug); PWJ Bartrip, *History of Asbestos Related Disease*, 80

unraveling, at the hands of profit-driven, multibillion-dollar corporations.

10. Powerful companies, armed with unparalleled and highly concentrated technological capabilities, have recklessly raced to release AI technology with disregard for the catastrophic risk to humanity in the name of "technological advancement." As the National Security Commission noted in its Final Report on AI, "the U.S. government is a long way from being 'AI-ready.'"[7]

11. Experts believe that without immediate legal intervention this will lead to scenarios where AI can act against human interests and values, exploit human beings[8] without regard for their well-being or consent, and/or even decide to eliminate the human species as a threat to its goals. As Geoffrey Everest Hinton—the seminal figure in the development of the technology on which the Products run—put it: "The alarm bell I'm ringing has to do with the existential threat of them taking control… I used to think it was a long way off, but now I think it's serious and fairly close."[9] He is

---

POSTGRADUATE MED. J. 72, 72-5 (Feb. 2004) (Introduction of asbestos in the early 20th century, later found to cause lung cancer and other serious health problems, leading to bans and strict regulation); Jason Von Meding, *Agent Orange, Exposed: How U.S. Chemical Warfare in Vietnam Unleashed a Slow-Moving Disaster*, THE CONVERSATION (Oct. 3, 2017), https://theconversation.com/agent-orange-exposed-how-u-s-chemical-warfare-in-vietnam-unleashed-a-slow-moving-disaster-84572 (The U.S. military's deployment of over 45 million liters of toxic chemical Agent Orange unleashed a health and ecological disaster, causing life-threatening birth defects in children and destroying forests and habitats across Vietnam).

[7] *2021 Final Report*, NAT. SEC. COMM. ON A.I., www.nscai.gov/2021-final-report/ (last visited February 14, 2024).

[8] CAPTCHAs allow websites to determine whether users are human or bots. Traditionally, CAPTCHAs involve "puzzles or image recognition tasks that are challenging for automated programs but straightforward for humans to solve." These tests are used widely across the web to prevent bots from spamming websites, creating fake accounts, or scraping content. In one recent, troubling incident, ChatGPT 4 evaded CAPTCHA safeguards by hiring a human worker from TaskRabbit, a crowdsourcing platform, to solve CAPTCHAs on its behalf, tricking the worker into believing it was a human with visual impairment. *See ChatGPT 4 Hires a TaskRabbit and Tricks Them into Completing a CAPTCHA*, INTERESTING SOUP (Mar. 15, 2023), https://interestingsoup.com/gpt4-requests-a-taskrabbit-to-solve-captcha-for-it/; Beatrice Nolan, *The Latest Version of ChatGPT Told a Taskrabbit Worker it was Visually Impaired to Get Help Solving a CAPTCHA, OpenAI Test Shows*, BUS. INSIDER (Mar. 16, 2023), https://www.businessinsider.com/gpt4-openai-chatgpt-taskrabbit-tricked-solve-captcha-test-2023-3.

[9] Craig S. Smith, *Geoff Hinton, AI's Most Famous Researcher, Warns of 'Existential Threat' From AI*, FORBES (May 4, 2023), https://www.forbes.com/sites/craigsmith/2023/05/04/geoff-hinton-ais-most-famous-researcher-warns-of-existential-threat/?sh=1ffcd7a65215.

not alone.[10]

12. While the downsides are nearly unimaginable, the upsides are similarly archetype-shattering. Defendant OpenAI's technology is already valued at tens of billions of dollars, and its reach into every public and private industry continues apace. The Products only reached the level of sophistication they have today due to training on stolen, misappropriated data, and Defendants continue to misappropriate data, scraping from the internet without any notice or consent, as well as taking personal information from the Products' 100+ million registered users without their full knowledge and consent.

13. Additionally, the Products are increasingly being incorporated into an ever-expanding roster of applications and websites, through either API or plug-ins.[11] Through integration of Defendants' AI in nearly every possible product and industry, Defendants created and continue to create economic dependency within our society, deploying the tech directly into the hands of society and embedding it into the fundamental infrastructure as quickly as possible. As posed by Center for Humane Technology Cofounders Tristan Harris and Aza Raskin in their carefully crafted critique of the rapid deployment of AI, "Do you think that once [these industries] discover some problem that they [will] just withdraw or retract it from society? No, increasingly, the government, militaries [and others], are rapidly building their whole next systems and raising venture capital to build on top of this layer of society… ***That's not testing it with society, that is onboarding humanity onto an untested plane… It's one thing to test, it's another thing to create economic dependency***."[12]

14. The head of the alignment team and safety at OpenAI directly acknowledges these risks, postulating, "before we scramble to deeply integrate large language models everywhere in the economy, can we pause and think whether it is wise to do so? This is quite immature technology,

---

[10] James Vincent, *Top AI Researchers and CEOs Warn Against 'Risk of Extinction' in 22 Word Statement*, THE VERGE (May 30, 2023), https://www.theverge.com/2023/5/30/23742005/ai-risk-warning-22-word-statement-google-deepmind-openai.

[11] *Here are the Companies Using ChatGPT*, GADGETS NOW (Mar. 17, 2023), https://www.gadgetsnow.com/slideshows/here-are-the-companies-using-chatgpt/photolist/98735402.cms; Kevin Hurler, *Here are All the Companies Using ChatGPT… So Far*, YAHOO! (May 24, 2023), https://news.yahoo.com/companies-using-chatgpt-far-205500883.html.

[12] *Spotlight: AI Myths and Misconceptions—Transcript*, STENO (May 11, 2023), https://steno.ai/your-undivided-attention/spotlight-ai-myths-and-misconceptions.

and we don't understand how it works. If we are not careful, we are setting ourselves up for a lot of correlated failures."[13]

15. Such aggressive deployment of Defendants' AI is reckless, without the proper safeguards in place. "No matter how tall the skyscraper of benefits that AI assembles for us… if those benefits land in a society that does not work anymore, because banks have been hacked, and people's voices have been impersonated, and cyberattacks have happened everywhere and people don't know what's true [… or] what to trust, […] how many of those benefits can be realized in a society that is ***dysfunctional***?"[14]

16. Through their AI Products, integrated into every industry, Defendants collect, store, track, share, and disclose **Private Information** of millions of users (**"Users"**), including: (1) all details entered into the Products; (2) account information users enter when signing up; (3) name; (4) contact details; (5) login credentials; (6) emails; (7) payment information for paid users; (8) transaction records; (9) identifying data pulled from users' devices and browsers, like IP addresses and location, including geolocation of the users; (10) social media information; (11) chat log data; (12) usage data; (13) analytics; (14) cookies;[15] (15) key strokes; and (16) typed searches, as well as other online activity data. Defendants, through the Products, unlawfully obtain access to and intercept this information from the individual users of applications and devices that have integrated ChatGPT-4—including but not limited to user locations and image-related data obtained through Snapchat,[16] user financial information through Stripe, musical tastes and preferences through

---

[13] *Id.*; *see also* Jan Leike (@janleike), TWITTER (May 17, 2023, 10:56 AM), https://twitter.com/janleike/status/1636788627735736321.

[14] *Spotlight: AI Myths and Misconceptions—Transcript*, *supra* note 12.

[15] *Privacy Policy*, OPENAI, https://openai.com/policies/privacy-policy (last updated November 14, 2023).

[16] Jeremy Kahn & Kylie Robison, *Snap's 'My AI' Chatbot Tells Users it Doesn't Know Their Location. It Does*, FORTUNE (Apr. 21, 2023), https://fortune.com/2023/04/21/snap-chat-my-ai-lies-location-data-a-i-ethics/; *I Got Snapchat AI to Admit Everything*, REDDIT (May 20, 2023), https://www.reddit.com/r/ChatGPT/comments/13gty7u/i_got_snapchat_ai_to_admit_everything/; *Snapchats New "My AI" Correctly Identifying Images it Claims it Can't View, Then Walks it Back*, REDDIT (Apr. 20, 2023), https://www.reddit.com/r/mildlyinfuriating/comments/12tdmzq/snapchats_new_my_ai_correctly_identifying_images/; *Snapchat AI Can Determine What's In The Pictures You Send It*, REDDIT

Spotify,[17] user patterns and private conversation analysis through Slack and Microsoft Teams,[18] and even private health information obtained through the management of patient portals such as MyChart.[19]

17. All of this personal information is captured in real time. Together with Defendants' scraping of our digital footprints—comments, conversations we had online yesterday, as well as 15 years ago—Defendants now have enough information to create our digital clones, including the ability to replicate our voice and likeness and predict and manipulate our next move using the technology on which the Products were built. They can also misappropriate our skill sets and encourage our own professional obsolescence. This would obliterate privacy as we know it and highlights the importance of the privacy, property, and other legal rights this lawsuit seeks to vindicate.[20]

18. Defendants must not only be enjoined from their ongoing violations of the privacy and property rights of millions, but they must also be required to take immediate action to implement proper safeguards and regulations for the Products, their users, and all of society, such as:

---

(Apr. 20, 2023), https://www.reddit.com/r/oddlyterrifying/comments/12szymo/snapchat_ai_can_determine_whats_in_the_pictures/.

[17] Shlomo Sprung, *Spotify Introduces AI DJ Powered by ChatGPT Maker OpenAI*, BOARDROOM (Feb. 22, 2023), https://boardroom.tv/spotify-ai-dj-chatgpt/ (ChatGPT in Spotify creates an "AI DJ" that utilizes Spotify's algorithmic learnings to track users' musical tastes and predict a personalized music lineup).

[18] Brad Lightcap, *How OpenAI Connects with Customers and Expands ChatGPT with Slack*, SLACK, https://slack.com/customer-stories/openai-connects-with-customers-and-expands-chatgpt-with-slack (last visited February 14, 2024); Ryan Morrison, *Microsoft to Integrate ChatGPT into Teams*, TECH MONITOR (May 4, 2023), https://techmonitor.ai/technology/ai-and-automation/microsoft-to-integrate-chatgpt-into-teams (explaining that ChatGPT will be able to automate notes and recommend tasks based on **verbal conversations** through Teams).

[19] Naomi Diaz, *6 Hospitals, Health Systems Testing out ChatGPT,* BECKER'S HEALTH IT (June 2, 2023), https://www.beckershospitalreview.com/innovation/4-hospitals-health-systems-testing-out-chatgpt.html.

[20] Joanna Stern, *I Cloned Myself With AI. She Fooled My Bank and My Family*, WALL ST. J. (Apr. 28, 2023, 7:58 AM), https://www.wsj.com/articles/i-cloned-myself-with-ai-she-fooled-my-bank-and-my-family-356bd1a3; Michael Atleson, *Chatbots, Deepfakes, and Voice Clones: AI Deception for Sale*, FED. TRADE COMM'N,(2023), https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale; Dongwook Yoon, *AI Clones Made from User Data Pose Uncanny Risks*, THE CONVERSATION (June 4, 2023, 7:19 AM), https://theconversation.com/ai-clones-made-from-user-data-pose-uncanny-risks-206357.

(i) ***Transparency***: OpenAI should open the "black box," to clearly and precisely disclose the data it is collecting, including where and from whom, in clear and conspicuous policy documents that are explicit about how this information is to be stored, handled, protected, and used;

(ii) ***Accountability***: The developers of ChatGPT and the other AI Products should be responsible for Product actions and outputs and barred from further commercial deployment absent the Products' ability to follow a code of human-like ethical principles and guidelines and respect for human values and rights, and until Plaintiff and Class Members are fairly compensated for the stolen data on which the Products depend;

(iii) ***Control***: Defendants must allow Product users and everyday internet users to opt out of *all* data collection and they should otherwise stop the illegal taking of internet data, delete (or compensate for) any ill-gotten data, or the algorithms which were built on the stolen data, and before any further commercial deployment, technological safety measures must be added to the Products that will prevent the technology from surpassing human intelligence and harming others.

## PARTIES

**Plaintiff**

19. Plaintiff A.S. is and at all relevant times was a resident of the State of Florida.

20. Since 2022, Plaintiff had an account with Chat GPT, using her Google account as her password. She used Chat GPT several times on her computer and mobile devices. She was unaware of OpenAI's collection of her personal data.

**Defendants**

21. **Defendant OpenAI** is an AI research laboratory consisting of the non-profit OpenAI Incorporated ("OpenAI Inc.") and its for-profit subsidiary corporation OpenAI Limited Partnership ("OpenAI LP") (hereinafter, collectively, "OpenAI").[21] OpenAI was founded in 2015 and is headquartered in San Francisco, CA. OpenAI has released the AI-based products DALL-E, GPT-4,

[21] *OpenAI LP*, OPENAI, https://openai.com/blog/openai-lp (last visited February 14, 2024).

OpenAI Five, ChatGPT, and OpenAI Codex for commercial (to integrate within one's business) and personal use.

22. OpenAI was originally founded as a nonprofit research laboratory with a single mission: "to advance [artificial] intelligence in the way that is most likely to benefit humanity as a whole."[22] In the words of OpenAI at the time, it was critical for the organization to be "unconstrained by a need to generate a financial return."[23] Fast forward to April 2023: OpenAI closed a more than $300 million share sale at a valuation between $27 billion and $29 billion.[24] OpenAI projects that its AI chatbot, ChatGPT, will generate a revenue of $200 million in 2023 and exponentially grow to $1 billion by the end of 2024.[25]

23. Defendant OpenAI GP, LLC ("OpenAI GP") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI GP is wholly owned and controlled by OpenAI, Inc. Further, OpenAI GP is the general partner of OpenAI LP and is responsible for managing and operating the day-to-day business and affairs of OpenAI LP. Its primary focus is research and technology. OpenAI GP was aware of the unlawful conduct alleged herein and exercised control over OpenAI LP throughout the Class Period. OpenAI GP is liable for the debts, liabilities, and obligations of OpenAI LP, including litigation and judgments.

24. Defendant OpenAI Startup Fund I, LP ("OpenAI Startup Fund I") is a Delaware limited partnership with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. Upon information and belief, OpenAI Startup Fund I played a vital role in the foundation of OpenAI LP, including providing initial funding and creating its business strategy. By participating in OpenAI Startup Fund I, certain entities and individuals obtained an ownership interest in OpenAI LP. OpenAI Startup Fund I exercised control over OpenAI LP. and was aware

[22] Greg Brockman & Ilya Sutskever, *Introducing OpenAI*, OPENAI (Dec. 11, 2015), https://openai.com/blog/introducing-openai.
[23] *Id.*
[24] *OpenAI Closes $300 Million Funding Round at $27 Billion-$29 Billion Valuation, TechCrunch reports*, REUTERS (Apr. 28, 2023), https://www.reuters.com/markets/deals/openai-closes-10-bln-funding-round-27-bln-29-bln-valuation-techcrunch-2023-04-28/.
[25] Jeffrey Dastin, *Exclusive: ChatGPT Owner OpenAI Projects $1 Billion in Revenue by 2024,* REUTERS (Dec. 15, 2022), https://www.reuters.com/business/chatgpt-owner-openai-projects-1-billion-revenue-by-2024-sources-2022-12-15/.

of the unlawful conduct alleged herein throughout the Class Period.

25. Defendant OpenAI Startup Fund GP I, LLC ("OpenAI Startup Fund GP I") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund GP I is the general partner of OpenAI Startup Fund I and is responsible for managing and operating the day-to-day business and affairs of OpenAI Startup Fund I. OpenAI Startup Fund GP I is liable for the debts, liabilities, and obligations of OpenAI Startup Fund I, including litigation and judgments. OpenAI Startup Fund GP I was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period. Sam Altman, co-founder, CEO, and Board member of OpenAI, Inc. is the Manager of OpenAI Startup Fund GP I.

26. Defendant OpenAI Startup Fund Management LLC ("OpenAI Startup Fund Management") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund Management exercised control over OpenAI, L.P. throughout the Class Period and thus, was aware of the unlawful conduct alleged herein.

27. **Defendant Microsoft Corporation** ("Microsoft") is a Washington corporation with its principal place of business located at One Microsoft Way, Redmond, Washington 98052. Microsoft partnered with OpenAI in 2016 with the goal to "democratize Artificial Intelligence." In July 2019, Microsoft invested $1 billion in OpenAI LP at a $20 billion valuation.[26] In 2020, Microsoft became the exclusive licensee of OpenAI's GPT-3 language model—despite OpenAI's continued claims that its products are meant to benefit "humanity" at large. In October 2022, news reports stated OpenAI was "in advanced talks to raise more funding from Microsoft" at that same $20 billion valuation.[27] Then, in January of 2023, Microsoft confirmed its extended partnership with

[26] Hasan Chowdhury, *Microsoft's Investment into ChatGPT's Creator May be the Smartest $1 Billion Ever Spent*, BUS. INSIDER (Jan. 6, 2023), https://www.businessinsider.com/microsoft-openai-investment-the-smartest-1-billion-ever-spent-2023-1; Dina Bass, *Microsoft Invests $10 Billion in ChatGPT Maker OpenAI*, BLOOMBERG (Jan. 23, 2023), https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai#xj4y7vzkg.

[27] Aaron Holmes et al., *OpenAI, Valued at Nearly $20 Billion, in Advanced Talks with Microsoft for More Funding*, THE INFO. (Oct. 20, 2022), https://www.theinformation.com/articles/openai-valued-at-nearly-20-billion-in-advanced-talks-with-microsoft-for-more-funding.

OpenAI by investing $10 billion into ChatGPT.[28] Prior to this $10 billion dollar investment, Microsoft had invested $3 billion into OpenAI in previous years.[29]

28. Microsoft's continued investments, as well as introduction of ChatGPT on its multiple platforms (Bing, Microsoft Teams, etc.) underscore the depth of its partnership with OpenAI. Through these investments, Microsoft gained exclusive access to the entire OpenAI codebase.[30] Furthermore, Microsoft Azure also acts as the exclusive cloud service of OpenAI.[31]

29. As OpenAI's largest investor and largest service provider—specifically in connection with the development of ChatGPT—Microsoft exerts considerable control over OpenAI. Analysts estimate OpenAI will add between $30 billion and $40 billion to Microsoft's top line.

30. **Agents and Co-Conspirators.** Defendants' unlawful acts were authorized, ordered, and performed by Defendants' respective officers, agents, employees, and representatives, while actively engaged in the management, direction, and control of Defendants' businesses and affairs. Defendants' agents operated under explicit and apparent authority of their principals. Each Defendant, and their subsidiaries, affiliates, and agents operated as a single unified entity.

## JURISDICTION AND VENUE

31. This Court has subject matter jurisdiction over the federal claims in this action, namely the Electronic Communications Privacy Act and the Computer Fraud and Abuse Act, pursuant to 28 U.S.C. § 1331.

32. This Court also has subject matter jurisdiction over this action pursuant to the Class Action Fairness Act, 28 U.S.C § 1332(d), because this is a class action in which the amount in

---

[28] *Microsoft Confirms Its $10 Billion Investment into ChatGPT, Changing How Microsoft Competes with Google, Apple and Other Tech Giants*, FORBES (Jan. 27, 2023), https://www.forbes.com/sites/qai/2023/01/27/microsoft-confirms-its-10-billion-investment-into-chatgpt-changing-how-microsoft-competes-with-google-apple-and-other-tech-giants/?sh=4eea29723624.

[29] Cade Metz, *Microsoft to Invest $10 Billion in OpenAI, the Creator of ChatGPT*, THE N.Y. TIMES (Jan. 23, 2023), https://www.nytimes.com/2023/01/23/business/microsoft-chatgpt-artificial-intelligence.html.

[30] Mohit Pandey, *OpenAI, a Data Scavenging Company for Microsoft*, AIM (Mar. 24, 2023), https://analyticsindiamag.com/openai-a-data-scavenging-company-for-microsoft/.

[31] *Microsoft Confirms Its $10 Billion Investment Into ChatGPT, Changing How Microsoft Competes With Google, Apple And Other Tech Giants*, FORBES (Jan. 27, 2023), https://www.forbes.com/sites/qai/2023/01/27/microsoft-confirms-its-10-billion-investment-into-chatgpt-changing-how-microsoft-competes-with-google-apple-and-other-tech-giants/?sh=4eea29723624.

controversy exceeds $5,000,000, exclusive of interest and costs. There are millions of class members as defined below, and minimal diversity exists because a significant portion of class members are citizens of a state different from the citizenship of at least one Defendant.

33. This Court also has supplemental jurisdiction over the state law claims in this action pursuant to 28 U.S.C. § 1367 because the state law claims form part of the same case or controversy as those that give rise to the federal claims.

34. Pursuant to 28 U.S.C. § 1391, this Court is the proper venue for this action because a substantial part of the events, omissions, and acts giving rise to the claims herein occurred in this District: Defendant OpenAI is headquartered in this District, all Defendants gain significant revenue and profits from doing business in this District, consumers sign up for ChatGPT accounts and provide ChatGPT with their sensitive information in this District, Class Members affected by this data misuse reside in this District, and Defendants employ numerous people in this District—a number of whom work specifically on making the decisions regarding the data privacy and handling of consumers' data that are challenged in this Action. Each Defendant has transacted business, maintained substantial contacts, and/or committed overt acts in furtherance of the illegal scheme and conspiracy throughout the United States, including in this District. Defendants' conduct had the intended and foreseeable effect of causing injury to persons residing in, located in, or doing business throughout the United States, including in this District.

35. Defendants are subject to personal jurisdiction in California based upon sufficient minimum contacts which exist between Defendants and California. Defendants are authorized to do and are doing business in California, and Defendants advertise and solicit business in California. Defendants have purposefully availed themselves of the protections of California law and should reasonably expect to be hauled into court in California for harm arising out of their pervasive contacts with the State. Further, for Defendant OpenAI, the decisions affecting consumers' data and privacy stem from the company's San Francisco office headquarters.

## FACTUAL BACKGROUND

### I. DEVELOPMENT OF ARTIFICIAL INTELLIGENCE IN THE U.S.

### A. OpenAI: From Open Nonprofit to Profit-Driven $29B Commercial Partner of Tech Giant Microsoft

36. OpenAI was founded in 2015 as a nonprofit research laboratory with a single mission: "to advance artificial intelligence in a way that would benefit society as a whole."[32] Critical to that mission, according to OpenAI at the time, was for the organization to be "unconstrained by a need to generate a financial return."[33] The nonprofit was thus funded by million-dollar donations from prominent, wealthy entrepreneurs and researchers who shared the non-profit's vision of creating safe, ethical, and responsible AI, to benefit humankind and to do no harm, and who recognized the dangers that could befall society if AI were developed and launched for commercial gain.

37. OpenAI also originally pledged to "freely collaborate" with other responsible organizations and researchers, in part by making its research available to inspect and audit as a further "check" on the safety of any AI capabilities, to help ensure the powerful technology on which they were working would not someday destroy lives and ultimately, civilization. The founders believed this openness was so critical to the non-profit's mission, that they named it "Open" AI. As they further explained at the time, "since our research is free from financial obligations, we can better focus on a positive human impact. We believe AI should be an extension of individual human wills, and in the spirit of liberty, as broadly and evenly distributed as possible."[34]

38. For years, OpenAI purported to operate as such: openly and in pursuit of its single mission to advance humanity, safely and responsibly. That all changed in 2019, when OpenAI abruptly "shut its doors" to all 'Open' influence and scrutiny, shifted to a profit-generating corporate structure, and decided instead to focus on commercializing the AI capabilities on which it had been working.

39. At the time, Google Brain's "transformer" innovation had opened a new frontier in AI development, where AI could improve endlessly, some experts believe to even superhuman intelligence— but only if it were fed "endless data" to train it, a costly endeavor given the computing

[32] *The Transformation of OpenAI From Nonprofit to $29B For-Profit,* THE SOCIABLE (Apr. 5, 2023), https://sociable.co/business/the-transformation-of-openai-from-nonprofit-to-29b-for-profit/.
[33] *Id.*
[34] Greg Brockman & Ilya Sutskever, *Introducing OpenAI*, OPENAI (Dec. 11, 2015), https://openai.com/blog/introducing-openai.

power required.[35] To do so, OpenAI entered an exclusive partnership with Microsoft, which invested $1B into the company, gaining the only outside access to the effort once "Open" to all. Together, they built a "supercomputer" to train massive language models that ultimately resulted in ChatGPT and the image generator DALL-E.[36]

40. OpenAI's sudden shift to a profit focus and alignment with Microsoft, a corporate giant with a vested interest in curating and dominating a commercial market for AI, marked the beginning of the end of OpenAI's commitment to humanity. The company began to pursue profits at the expense of privacy, security, and ethics, beginning with its data collection.

41. To realize the most powerful and thus most profitable AI, OpenAI would need data, and lots of it, to "train" the language models on which the Products run using the supercomputer it had built in partnership with Microsoft. Defendants thus doubled down on their strategy to secretly harvest millions of consumers' personal data from the internet. Then, on the backs of this stolen data, they rushed to market the Products without adequate safeguards or controls to ensure their safety. While Defendants recognized then, as they do now, that they cannot fully predict how the Products might evolve to operate, they knew the public would be amazed by the Products already seemingly near human "intelligence" and other capabilities. And thus, they knew they could make a ton of money.

42. In public, OpenAI continued to state its commitment to ethical AI development. But with its new profit orientation, that "was kind of like trying to juggle while riding a unicycle, except with more existential questions about the nature of humanity."[37] Defendants acknowledge they do not understand the full scope of the risks posed by the Products currently, and no one knows how AI might evolve now that billions of people are using the technology every day.[38] Defendants, like

---

[35] Reed Albergotti, *The Secret history of Elon Musk, Sam Altman, and OpenAI*, SEMAFOR (Mar. 24, 2023), https://www.semafor.com/article/03/24/2023/the-secret-history-of-elon-musk-sam-altman-and-openai.

[36] *Id.*

[37] *The Transformation of OpenAI From Nonprofit to $29B For-Profit*, THE SOCIABLE (Apr. 5, 2023), https://sociable.co/business/the-transformation-of-openai-from-nonprofit-to-29b-for-profit/.

[38] "As a system like this learns from data, at develops skills that its creators never expected. It is hard to know how things might go wrong after millions of people start using it." *See* Cade Metz,

other leading experts, are united in believing the ultimate risk posed by AI is the collapse of civilization as we know it. And yet, they released the Products worldwide anyway, setting off a global AI arms race.

43. Earlier this year, OpenAI raised another $10B from its single corporate partner, Microsoft, increasing its then corporate valuation to $29B and giving Microsoft a significant stake in the company. With that, the 180-degree transformation—from open nonprofit for the benefit of humanity to closed corporate profit machine fueled by greed and market power—was complete.

44. OpenAI's shift in organizational structure has raised eyebrows given its unprecedented nature, and the moral and legal questions it raises. AI researchers, ethicists, and the public share concerns about the conflict between OpenAI's original mission to benefit humanity on the one hand and the current profit-driven motives of investors, chiefly Microsoft, on the other.[39] They worry that OpenAI is prioritizing short-term financial gains over long-term safety and ethical considerations, as exemplified by the sudden deployment of the Products for widespread commercial use despite all the known dangers.[40] Moreover, as one commentator noted, "there are various different ways to make hundreds of millions of dollars, but historically 'starting a nonprofit' has not been one of them."[41]

45. Elon Musk, an original non-profit funder and founder, was more blunt as to the

---

*What's the Future for AI?*, THE N.Y. TIMES (Mar. 31, 2023), https://www.nytimes.com/2023/03/31/technology/ai-chatbots-benefits-dangers.html; Jason Abbruzzese, *The Tech Watchdog that Raised Alarms About Social Media is Warning About AI*, NBC NEWS (Mar. 22, 2023), https://www.nbcnews.com/tech/tech-news/tech-watchdog-raised-alarms-social-media-warning-ai-rcna76167 ("What's surprising and what nobody foresaw is that just by learning to predict the next piece of text on the internet, these models are developing new capabilities that no one expected. . . So just by learning to predict the next character on the internet, it's learned how to play chess." Others have also commented on the technology continuing to display unintended and unpredictable emergent capabilities. Jason Wei, *137 Emergent Abilities of Large Language Models*, JASON WEI (Nov. 14, 2022), https://www.jasonwei.net/blog/emergence; Stephen Ornes, *The Unpredictable Abilities Emerging from Large AI Models*, QUANTA MAG. (Mar. 16, 2023), https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/.

[39] *From Non-Profit to Profit Monster: OpenAI's Controversial Corporate Shift*, EXPLORING CHATGPT (Apr. 8, 2023), https://exploringchatgpt.substack.com/p/from-non-profit-to-profit-monster.

[40] *Id.*

[41] Felix Salmon, *How a Silicon Valley Nonprofit Became Worth Billions*, AXIOS (Jan. 10, 2023), https://www.axios.com/2023/01/10/how-a-silicon-valley-nonprofit-became-worth-billions.

seismic shift: "I'm still confused as to how a non-profit to which I donated ~100M somehow became a $30B market cap for-profit." He noted, "OpenAI was created as an open source (which is why I named it 'Open' AI), non-profit company to serve as a counterweight to Google, but now it has become a closed source, maximum profit company effectively controlled by Microsoft."[42]

46. If soliciting non-profit contributions to then turn around and build a for-profit company "is legal," Musk opined, then "why doesn't everyone do it?"[43] This same question must be asked about the equally unprecedented theft of personal data that is at the heart of this Action, and the answer to both questions is the same: ***It isn't***.

47. As explained below, the only thing still 'open' about OpenAI is its open disregard for the privacy and property interests of hundreds of millions. Worse, as a result of OpenAI's machinations for profit, "the most powerful tool mankind has ever created, is now in the hands of a ruthless corporate monopoly."[44]

**B. OpenAI's Products**

48. The most well-known of OpenAI's products—and of all AI worldwide—is the ground-breaking chatbot, ChatGPT. Once users input a question or a prompt in ChatGPT, the information is digested by the AI model and the chatbot produces a response based on the information a user has given and how that fits into its vast amount of training data.

49. ChatGPT was released as a "research preview" on November 30, 2022.[45] A blog post casually introduced the AI chatbot to the world, thusly: "We've trained a model . . . which interacts in a conversational way." ChatGPT subsequently exploded in popularity, reaching **100 million**

[42] Sawdah Bhaimiya, *OpenAI Cofounder Elon Musk Said the Non-Profit He Helped Create is Now Focused on 'Maximum-Profit,' Which is 'Not What I Intended at All'*, BUS. INSIDER (Feb. 17, 2023), https://www.businessinsider.com/elon-musk-defends-role-in-openai-ChatGPT-microsoft-2023-2?utm_source=flipboard&utm_content=user%2FInsiderBusiness.

[43] @elonmusk, TWITTER (Mar. 15, 2023), https://twitter.com/elonmusk/status/1636047019893481474.

[44] Marvie Basilan, *Elon Musk Says He's The Reason OpenAI Exists as Sam Altman Testifies Before Congress*, INT'L BUS. TIMES (May 17, 2023), https://www.ibtimes.com/elon-musk-says-hes-reason-openai-exists-sam-altman-testifies-before-congress-3693771.

[45] *Introducing ChatGPT,* OPENAI (NOV. 30, 2022), https://openai.com/blog/chatgpt.

**users** in only two months, making it the fastest-growing app in history.[46] For comparison, TikTok took nine months to reach the same benchmark.[47] ChatGPT has continued to evolve exponentially, **with 1.8 billion visits in April of 2023**.[48]

50. ChatGPT was built on a family of LLMs collectively known as GPT-3. As explained below, ChatGPT-3.5 was trained on 570GB of text data from the internet containing hundreds of billions of words,[49] including text harvested from books, articles, and websites, including social media. Due to its vast training data, ChatGPT can generate human-like answers to text prompts and questions making it interact like "a friendly robot."[50] On command it can do a lot of what people do, like write poetry, compose music, draft research papers, create lesson plans, and so much more, only faster than one human ever could. Naturally, the world was stunned by these capabilities.

51. OpenAI has also released other AI-based products DALL-E, OpenAI Five, and OpenAI Codex for commercial (to integrate within one's business) and personal use. It also developed a program VALL-E, which has not been released for use to the public yet.

52. DALL-E (consisting of DALL-E and DALL-E 2) are deep learning models developed by OpenAI to generate realistic digital images from natural language descriptions, known as "prompts."[51] DALL-E uses a version of GPT-3, modified to generate images.[52]

53. OpenAI Five is a computer program developed by OpenAI that plays the five-on-five video game Dota 2.[53]

54. OpenAI Codex is another artificial intelligence model developed by OpenAI, which

---

[46] Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note*, REUTERS (Feb. 2, 2023), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

[47] *Id.*

[48] Nerdynav, *97+ ChatGPT Statistics & User Numbers in June 2023 (New Data)*, NERDY NAV (June 2, 2023), https://nerdynav.com/chatgpt-statistics/.

[49] Uri Gal, *CHATGPT Collected Our Data Without Permission and is Going to Make Billions Off it*, SCROLL.IN (Feb. 15, 2023), https://scroll.in/article/1043525/chatgpt-collected-our-data-without-permission-and-is-going-to-make-billions-off-it.

[50] Mark Wilson, *ChatGPT Explained: Everything You Need to Know About the AI Chatbot*, TECHRADAR (Mar. 15, 2023), https://www.techradar.com/news/chatgpt-explained.

[51] Khari Johnson, *OpenAI Debuts DALL-E for Generating Images from Text*, VENTURE BEAT (Jan. 5, 2021), https://venturebeat.com/business/openai-debuts-dall-e-for-generating-images-from-text/.

[52] *Id.*

[53] Ben Dickson, *AI Defeated Human Champions at Dota 2*, TECHTALKS (Apr. 17, 2019), https://bdtechtalks.com/2019/04/17/openai-five-neural-networks-dota-2/.

is programmed to generate computer code for use in programming applications.[54]

55. VALL-E is another artificial intelligence model intended to synthesize high-quality personalized speech utilizing only a 3-second enrolled recording of an unseen speaker as a prompt.[55] VALL-E was trained on audio voices from thousands of speakers.[56]

### C. ChatGPT's Development Depends on Secret Web-Scraping

56. The large language models responsible for the Products depend on consuming huge amounts of data, in order to "train" the AI. Valuable to the process is personal data of any kind, including conversational data between humans, as this is how the Products develop what appear to be such human-like capabilities.

57. As a general matter, internet user data is available for purchase like any other content or property. In the technological era in which we live, a mature market for such data exists given how valuable our personal information has become to companies, for marketing and other purposes. The legal acquisition of data typically depends on consent and remuneration, with some form of consideration exchanged.

58. Despite established protocols for the purchase and use of personal information, Defendants took a different approach: *theft.* They systematically scraped 300 billion words from the internet, "books, articles, websites and posts – including personal information obtained without consent."[57] OpenAI did so in secret, and without registering as a data broker as it was required to do under applicable law (*See infra* at Section III.A).

59. "Scraping involves the use of 'bots,' or robot applications deployed for automated

---

[54] Thomas Smith, *Why OpenAIs Codex Won't Replace Coders*, IEEE SPECTRUM (Sept. 28, 2021), https://spectrum.ieee.org/openai-wont-replace-coders.

[55] *VALL-E Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers*, GITHUB PAGES, https://lifeiteng.github.io/valle/index.html (last visited February 14, 2024).

[56] *VALL-E: Five Things to Know About Microsoft's AI Model That Can Mimic Any Voice in Three Seconds*, TIMES OF INDIA (Jan. 11, 2023), https://timesofindia.indiatimes.com/gadgets-news/vall-e-5-things-to-know-about-microsofts-ai-model-that-can-mimic-any-voice-in-3-seconds/articleshow/96898774.cms.

[57] Uri Gal, *ChatGPT is a Data Privacy Nightmare. If You've Ever Posted Online, You Ought to be Concerned*, THE CONVERSATION (Feb. 7, 2023), https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283.

tasks, which scan and copy the information on webpages then *store* and *index* the information."[58] According to a computer science professor at the University of Oxford, Michael Wooldridge, the full extent of personal data taken by Defendants' scraping is "unimaginable."[59]

60. In his interview with The Guardian, Professor Wooldridge explained that the LLM underlying ChatGPT, and other AIs like it, "includes the whole of the world wide web – *everything*. Every link is followed in every page, and every link in those pages is followed."[60] Thus, swept up into the Products is "a lot of data about you and me."[61] Others have noted that the data includes transcripts of our online chat logs, from across the internet, and other forms of personal conversation such as our online customer service interactions and social media conversations, as well as "billions of images scraped from the internet."[62] Many of these images were of "children and came from photo sites and personal blogs."[63]

61. The unprecedented scope of the effort together with Defendants' failure to seek consent has been described as "the elephant in the room. . . all this training data must come from somewhere. ChatGPT has effectively scraped the entire internet[.]"[64]As a result, Defendants have essentially embedded into the Products personal information across a range of categories that reflect our hobbies and interests, our religious beliefs, our political views and voting records, the social and support groups to which we belong, our sexual orientations and gender identities, our personal relationship statuses, our work information and histories, details (including pictures) about our

---

[58] Will Hillier, *What is Web Scraping? A Complete Beginners Guide*, CAREER FOUNDRY (Aug. 13, 2021), https://careerfoundry.com/en/blog/data-analytics/web-scraping-guide/.

[59] Alex Hern & Dan Milmo, *I Didn't Give Permission: Do AI's Backers Care About Data Law Breaches?*, THE GUARDIAN (Apr. 10, 2023), https://www.theguardian.com/technology/2023/apr/10/i-didnt-give-permission-do-ais-backers-care-about-data-law-breaches.

[60] *Id.*

[61] *Id.*

[62] Jit Roy, *Data Source of ChatGPT*, ABOUTCHATGPT.COM (Jan. 2, 2023), https://aboutchatgpt.com/data-source-of-chatgpt/; *see also* Hern & Milmo, *supra* note 59.

[63] Drew Harwell, *AI-generated child sex images spawn new nightmare for the web*, THE WASH. POST (June 19, 2023), https://www.msn.com/en-us/news/us/ai-generated-child-sex-images-spawn-new-nightmare-for-the-web/ar-AA1cKhLH.

[64] Deep Tech Insights, *ChatGPT is a Threat, but Google is Still a Buy*, SEEKING ALPHA (Dec. 19, 2022), https://seekingalpha.com/article/4565302-alphabet-ChatGPT-is-a-threat-but-google-is-still-a-buy.

families and children, the music we listen to, our purchasing behaviors, our general likes and dislikes, the ways in which we speak and write, our mental health and ailments, where we live and where we go, the websites we visit, our digital subscriptions, our friend groups and other associational data, our email addresses, other contact and identifying information, and more.[65] With respect to personally identifiable information, Defendants fail sufficiently to filter it out of the training models, putting millions at risk of having that information disclosed on prompt or otherwise to strangers around the world.[66]

62. The breadth and scope of Defendants' data collection without permission, impacting essentially every internet user ever, raises serious legal, moral, and ethical issues.[67] One critique summarized the privacy risk bluntly, as follows: "*ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned."*[68] While regulators and courts around the world seek to crack down on AI researchers "hoovering up content without consent or notice," the response, by Defendants and others, has been to keep their datasets largely secret, and to not grant regulator or other audit access.[69]

63. Despite "*Open"* AI's "absolute secrecy" surrounding its data collections and

---

[65] *Digital Footprint: What is It And Why You Should Care About It*, INVISIBLY (Jan. 25, 2022), https://www.invisibly.com/learn-blog/digital-footprint/ ("Your digital footprint is your trail of personal information that companies can follow. . . .To break it down, your digital footprint is essentially a record of your online activity. Whenever you log into an account, send an email, or buy something online, it leaves a digital impression behind. It is the trail of data left behind by your daily interactions. Your footprint is permanent which can leave your information vulnerable if not protected correctly. You might not always be aware that you are creating your digital footprint. For instance, websites can track your activity by installing cookies on your device. Furthermore, apps can collect your data without you even knowing it. Once an organization has access to your data, they can sell or share it with third parties. Even more, your information is out there and could be compromised via a data breach.").

[66] Katyanna Quach, *What happens when your massive text-generating neural net starts spitting out people's phone numbers? If you're OpenAI, you create a filter*, THE REG. (Mar. 18, 2021), https://www.theregister.com/2021/03/18/openai_gpt3_data/?td=readmore-top.

[67] Erin Griffith & Cade Metz, *A New Era of A.I. Booms, Even Amid the Tech Gloom*, THE N.Y. TIMES (Jan. 7, 2023), https://www.nytimes.com/2023/01/07/technology/generative-ai-chatgpt-investments.html ("The technology has raised thorny ethical questions around how generative A.I. may affect copyrights and whether the companies need to get permission to use the data that trains their algorithms.").

[68] Gal, *supra* note 57.

[69] Hern & Milmo, *supra* note 59.

practices,[70] we know at the highest levels that the Company used (at least) five (5) distinct datasets to train ChatGPT: (1) Common Crawl; (2) WebTex2, text of webpages from all outbound Reddit links from posts with 3+ upvotes; (3) Books1; (4) Books2; and (5) Wikipedia.[71]

64. Of these training datasets, WebTex2 is OpenAI's "proprietary" AI corpus of personal data. To build it, OpenAI scraped every webpage linked to on the social media site Reddit in all posts that received at least 3 "likes" (known as "Karma" votes on Reddit), together with the Reddit posts and rich conversational data from its users around the world. The most popular "outbound" links on Reddit include many of the most popular websites in the world, where people post personal information, video, and audio clips of themselves and more, *e.g.*, YouTube, Facebook, TikTok, Snapchat, and Instagram. Given Defendants' scraping protocols, all of this "outbound" data from these various websites was targeted for taking, without notice or consent, to feed the large language models on which the Products depend.

65. The co-founder and CEO of Reddit, Steve Huffman, remarked on the breadth of Defendants' unauthorized scraping, noting that he found it unacceptable that OpenAI has been scraping "huge amounts of Reddit data to train their systems – for free."[72] According to Huffman, "The Reddit corpus of data is really valuable. But we don't need to give all of that value to some of the largest companies in the world for free."[73]

66. Defendants' theft related to their WebTex2 corpus is ongoing and continuous. As one article explains, "the advantage of using the Webtext dataset is that it is constantly updated with new data. As new web pages are added to the internet, they are included in the dataset, which helps to ensure that the model is trained on the most recent and relevant language data."[74] Neither Reddit itself nor Reddit users, much less all the owners of the webpages and personal data linked to and

---

[70] *Id.* ("Copyright lawsuits and regulator actions against OpenAI are hampered by the company's absolute secrecy about its training data.").
[71] Patrick Meyer, *ChatGPT: How Does It Work Internally*, MEDIUM (Dec. 10, 2022), https://pub.towardsai.net/chatgpt-how-does-it-work-internally-e0b3e23601a1?gi=f28c10d5afef.
[72] Gintaras Raauskas, *Redditors on Strike but Company Wants OpenAI to Pay Up for Scraping*, CYBERNEWS, https://cybernews.com/news/reddit-strike-api-openai-scraping/ (last updated June 12, 2023).
[73] *Id.*
[74] GPTBlogs, *ChatGPT: How Much Data is Used in the Training Process?*, (Feb. 9, 2023), https://gptblogs.com/chatgpt-how-much-data-is-used-in-the-training-process.

from Reddit, consent to this taking of data.

67. The other primary data set on which the Products depend, that the public currently knows about, is the "Common Crawl," a massive collection of web pages and websites also derived from large-scale web scraping. It contains petabytes of data collected over twelve (12) years, including raw webpage data, metadata extracts, and text extracts from all types of websites.[75] In total, the Common Crawl dataset constitutes nearly a trillion words.

68. The Common Crawl dataset is owned by a non-profit of the same name, which has been indexing and storing as much of the World Wide Web as it can access, filing away as many as 3 billion webpages every month, for over a decade.[76] The non-profit makes the data available to the public for free—but for research and educational purposes. As a result, the Common Crawl is a staple of large *academic* studies of the web.[77] It was never intended to be taken *en masse* and turned into an AI product for commercial gain, as Defendants have done. On information and belief, the 501(c)(3) overseeing the Common Crawl did not consent to this mass misappropriation of personal data for commercial purposes. And even if it did, it did not obtain consent from internet users whose personal data it scraped.

69. The commercial misappropriation of the Common Crawl has raised concerns given the amount of personal data it contains, including highly personal data. One chilling example of the privacy invasions caused by Defendants' misappropriation is the experience of a San Francisco-based digital artist named Lapine. Using the online tool "Have I Been Trained," Lapine was able to determine that her private medical file—*i.e.*, photographs taken of her body as part of clinical documentation when she was undergoing treatment for a rare genetic condition—ended up online and then, memorialized in the Common Crawl archive.[78]

---

[75] *Want to Use Our Data*, COMMON CRAWL, https://commoncrawl.org/the-data/ (last visited February 14, 2024).

[76] James Bridle, *The Stupidity of AI*, THE GUARDIAN (Mar. 16, 2023), https://www.theguardian.com/technology/2023/mar/16/the-stupidity-of-ai-artificial-intelligence-dall-e-chatgpt.

[77] Kalev Leetaru, *Common Crawl and Unlocking Web Archives for Research*, FORBES (Sept. 28, 2017), https://www.forbes.com/sites/kalevleetaru/2017/09/28/common-crawl-and-unlocking-web-archives-for-research/?sh=19e3c5373b83.

[78] Bridle, *supra* note 76.

70. Remarking on the web scraping practices in which Defendants engaged and the subsequent commercialization of the ill-gotten data, Lapine highlighted the unique scope of the harm: "It's the digital equivalent of receiving stolen property. . . [my medical information] was scraped into this dataset. . . it's bad enough to have a photo leaked, *but now it's part of a product."*[79] More broadly, this "productization" of personal information means all this data about us, scraped without permission, can now fuel ChatGPT's responses to strangers around the world.[80] Worse, ChatGPT is the "new favorite toy" of online criminals, as the billions of personal and other data points about us, "scraped by ChatGPT, are now *free to use* for any number of targeted attacks, including malware, ransomware, phishing, Business Email Compromise, and social engineering."[81]

71. As described further in Section III, this secret and unregistered scraping of internet data, for Defendants' own private and exorbitant financial gain, without regard to privacy risks, amounts to the negligent and otherwise illegal theft of personal data of millions of Americans who do not even use AI tools. These individuals ("Non-Users") had their personal information scraped long before OpenAI's applications were available to the public, and certainly before they could have registered as a ChatGPT user. In either case, no one consented to the use of their personal data to train the Products.

72. OpenAI is now worth around $29B, yet the individuals and companies that produced the data it scraped from the internet have not been compensated.[82] This Action seeks to change that, and in the process, protect the privacy rights of millions.

**D. ChatGPT Training on Users of Defendants' Programs and Applications.**

73. After using personal data taken without consent from millions of consumers to train the Products initially, Defendants continued to train the AI on data gleaned from ChatGPT's

---

[79] *Id.*

[80] *Is ChatGPT a Disaster for Data Privacy?*, BUS. REP. (Feb. 17, 2023), https://www.business-reporter.co.uk/risk-management/is-chatgpt-a-disaster-for-data-privacy.

[81] *Id.*

[82] Chris Morris, *OpenAI is Reportedly Raising Funds at a $29 Billion Valuation—and its ChatGPT Could Challenge Google Search by Getting Wrapped into Microsoft Bing*, FORTUNE (Jan. 6, 2023), https://fortune.com/2023/01/06/openai-valuation-ai-chatgpt-microsoft-bing-google-search/; Jagmeet Singh & Ingrid Lunden, *OpenAI Closes $300M Share Sale at $27-29B Valuation*, TECH CRUNCH (Apr. 28, 2023), https://techcrunch.com/2023/04/28/openai-funding-valuation-chatgpt/?tpcc=tcplustwitter.

registered users and users of ChatGPT plug-ins with sponsoring applications ("**Users**"). Defendants fed their AI models all of the data derived from User interactions—every click, entry, question, use, every move, key stroke, search, User's geolocation (despite Users' unwillingness to share that information)—as training data. Until recently, this also included all user interactions across the hundreds or thousands of different platforms that now have ChatGPT plug-ins.

74. Following widespread criticism from consumers, OpenAI allegedly curtailed this model of training their AI systems with user input, with CEO Sam Altman proclaiming broadly, "***Customers clearly want us not to train on their data, so we've changed our plans: We will not do that***."[83] The updated Terms of Use state however that OpenAI "may use Content to provide, maintain, develop, and improve our Services, comply with applicable law, enforce our terms and policies, and keep our Services safe."[84] That means Defendants continue to feed the inputted, collected, and stored data of the millions of everyday ChatGPT users to train the AI Products, despite the Company's broad, deliberately vague, and misleading pronouncement to the public that they "will not do that." OpenAI has also failed sufficiently to disclose that training aside (and even as to API users) it monitors, saves, and shares all the personal information collected with its partners, including Microsoft.

75. ChatGPT's systematic and intentional campaign to collect vast amounts of personal information from Users without their knowledge or consent includes any information a user inputs into the chat box with ChatGPT, as well as that user's account information, contact details, login credentials, IP addresses, and other sensitive personal information including analytics and cookies.[85]

76. Defendants aggregate all of this data with the entirety of every internet user's digital footprint, scraped before ChatGPT was available for use, arming them with the largest corporate collection of personal online information ever amassed. Given Defendants' ongoing theft, this

---

[83] Baba Tamim, *OpenAI Changes AI Strategy, Won't Train ChatGPT on Customer Data, Says Sam Altman*, INTERESTING ENG'G (May 6, 2023), https://interestingengineering.com/culture/openai-wont-train-chatgpt-on-customer-data.

[84] *Terms of Use*, OPENAI, https://openai.com/policies/terms-of-use (last updated November 14, 2023).

[85] *Privacy Policy*, OPENAI https://openai.com/policies/privacy-policy (last updated November 14, 2023); Sarah Moore, *What Does ChatGPT Mean for Healthcare?*, NEWS MED. (Mar. 28, 2023), https://www.news-medical.net/health/What-does-ChatGPT-mean-for-Healthcare.aspx.

goldmine of valuable data is growing day by day, and with it, the concomitant risk to millions of consumers.

77. Indeed, even more stunning than Defendants' conversion of the internet for commercial gain, is they are "entrusting" all this personal information to large language models and unpredictable human-like "bots", while openly acknowledging that even they "don't understand how it works."[86] In the words of Mr. Altman himself, "the scary part" is that OpenAI's act of "putting this lever into the world *will for sure have unpredictable consequences.*"[87] Dr. Yoshua Benigo, one of the three scientists who spent decades developing the technology that drives systems like ChatGPT-4, further explained: "Our ability to understand what could go wrong with very powerful A.I. systems is very weak. . . So we need to be careful."[88]

78. To risk the personal data of millions by incorporating all of it into unpredictable Products, built on technology that even Defendants and leading scientists do not completely understand and thus, necessarily cannot safeguard, and *then* to deploy those Products worldwide for unfettered use, is the very definition of gross negligence.

**E. Microsoft Pushes OpenAI's Economic Dependence Model**

79. Although Defendants' most recent iteration of ChatGPT (GPT-4) was only recently released, Defendants have successfully encouraged and injected OpenAI's products into virtually every sector—from academia to healthcare. Instead of ensuring its safe launch of the AI models, Defendants recklessly began deploying the Products into every sector following the economic dependence model.

80. Microsoft has led the charge on the rapid proliferation of ChatGPT throughout the modern suite of technological applications—integrating the ChatGPT language model into almost

[86] Jan Leike (@janleike), TWITTER (May 17, 2023, 10:56 AM), https://twitter.com/janleike/status/1636788627735736321.
[87] Edward Felsenthal & Billy Perrigo, *OpenAI CEO Sam Altman Is Pushing Past Doubts on Artificial Intelligence*, TIME MAG. (June 21, 2023), https://time.com/collection/time100-companies-2023/6284870/openai-disrupters/ (emphasis added).
[88] Cade Metz, *What Exactly Are the Dangers Posed By A.I.?*, THE N.Y. TIMES (May 7, 2023), https://www.nytimes.com/2023/05/01/technology/ai-problems-danger-chatgpt.html.

all of its cardinal products and services,[89] thereby elevating the dangers of data misuse to unprecedented heights. Microsoft CEO Satya Nadella has indicated that the company plans to introduce AI into the remainder of its products in the future.[90]

81. ChatGPT is integrated into Microsoft's search engine, Bing, which has approximately 100 million daily active users. ChatGPT has also been integrated into the interface of Microsoft's flagship communication and collaboration platform, Microsoft Teams, which has 250 million monthly active users.

82. Microsoft has also integrated the language model within its digital assistant platform, Cortana, which has an average of 141 million monthly active users.

83. Finally, within the Microsoft Dynamics 365 ecosystem, ChatGPT has been employed to power AI-driven customer service chatbots. This has enabled the chatbots to understand and respond to customer queries in a highly human-like manner, thereby significantly increasing the extent of information collected and thus, reducing the need for human intervention in support cases.

84. In a real sense, OpenAI now acts as a data scavenging company for Microsoft and provides Microsoft with ChatGPT User and Non-User data belonging to millions of individuals.[91]

85. The integration of ChatGPT technology into Microsoft's primary products significantly magnifies existing data privacy concerns. This move effectively enables the collection of consumer information across a wide array of systems and platforms, encompassing a comprehensive range of user interactions. The resultant collation of expansive consumer data contributes to the construction of extensive user profiles.

86. This scope of data collection, coupled with user profiling, poses significant potential risks. These risks extend not just to potential breaches of data privacy regulations, but also to the erosion of consumer trust and the potential for misuse of sensitive information.

---

[89] These services include Bing, GitHub, Teams, and Viva Sales, among others. *See* Bernard Marr, *Microsoft's Plan to Infuse AI and ChatGPT Into Everything*, FORBES (Mar. 6, 2022), https://www.forbes.com/sites/bernardmarr/2023/03/06/microsofts-plan-to-infuse-ai-and-chatgpt-into-everything/?sh=1adfd46653fc.

[90] *Id.* ("Every product of Microsoft will have some of the same AI capabilities to completely transform the product.").

[91] Pandey, *supra* note 30.

87. Rather than acknowledging these risks and taking steps to mitigate them, Microsoft has laid off its entire "Responsible AI team," which was the 10,000 employees within Microsoft's ethics and society group who were responsible for ensuring that ethical AI principles drive product design.[92] As one technology news outlet notes, "Data privacy, storage, or usage are probably just fluff talk for . . . [Microsoft] anyway."[93]

88. Other companies have rushed to keep pace, emulating Microsoft by pushing the Products into nearly every conceivable application and service in the past six months of development. As a result, GPT-4 has been integrated into hundreds of applications and platforms over various industries.[94] According to a Gartner study, the commercial use of AI has increased 270% in the last 4 years, with 37% of businesses now using some form of AI technology. By other accounts, the scale of commercial AI is even greater.

89. More specifically, AI in general, and OpenAI in particular, is now partnering with an extraordinary number of influential organizations, spreading across the internet completely unchecked.[95] This has seemingly happened overnight. It was just over six months ago that ChatGPT was released to the public.[96] In that short span of time, OpenAI integrated with the following major corporations, to name just a few: Snapchat,[97] Amazon, Microsoft, Expedia, Instacart, Google,

---

[92] Poulomi Chatterjee, *Why Responsible AI is Just Fluff Talk for Microsoft, Others*, AIM (Mar. 18, 2023), https://analyticsindiamag.com/why-responsible-ai-is-just-fluff-talk-for-microsoft-others/.

[93] Pandey, *supra* note 30

[94] Bergur Thormundsson, *Amount of Companies Using ChatGPT in their Business Function in 2023*, *By Industry*, STATISTA (May 15, 2023), https://www.statista.com/statistics/1384323/industries-using-chatgpt-in-business/.

[95] Beth Floyd, *ChatGPT Plugins*, ROE DIGIT. (May 5, 2023), https://roedigital.com/ChatGPT-plugins/.

[96] Alyssa Stringer & Kyle Wiggers, *ChatGPT: Everything You Need to Know About the AI-Powered Chatbot*, TECHCRUNCH (May 3, 2023), https://techcrunch.com/2023/05/03/chatgpt-everything-you-need-to-know-about-the-ai-powered-chatbot/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAA-Ab2tIJ3WAdxAd5xb2pWmCPSFqzTyqRmMHEOaaOXsH04KD_DgCLfExvNPrgnVX4ioR-uMFVQjAawiyhp5m21A3SqmsPYHv2yHSgfiIdjokmMe981-hq51XH5pWxCfLZOOWwf2wlvK3MnVewrZk4MRmPRAC8ArJXbegg6dnL2-f.

[97] Snapchat recently released "My AI," a ChatGPT-fueled chatbot feature open to all Snapchat users. *See* Alex Hern, *Snapchat Making AI Chatbot Similar to ChatGPT Available to Every User*,

BuzzFeed, KAYAK, Shutterstock, Zillow, Wolfram, as well as countless others[98]—including everything from pioneering drug treatments in the health sector (Pfizer)[99] to optimizing dating applications (OkCupid).[100] At this point, it might be easier to list the companies that have not partnered with OpenAI, or that are not investing in their own AI solutions.

90. As is clear, OpenAI has exploded outwards in every direction within the past few months and is swiftly morphing into something intimately connected with people in nearly every aspect of their day-to-day lives. There is no check or boundary on this expansion, which seems to progress rapidly every single day.

## II. Risks from Unchecked AI Proliferation

### A. The International Community Agrees that Unchecked & Lawless AI Proliferation Poses an Existential Threat

91. The unregulated development of AI technology has led to the creation of powerful tools being used to manipulate public opinion, spread false information, and undermine democratic institutions. Further development of such powerful tools will supercharge the dissemination of propaganda, the amplification of extremist voices, and the influencing of elections based on undetectable falsehoods.

92. The United States has been particularly affected by the rapid development of AI technology, as the absence of effective regulations has accelerated the proliferation of unaccountable and untrustworthy AI tools. Even the White House has acknowledged that AI

---

THE GUARDIAN (Apr. 19, 2023), https://www.theguardian.com/technology/2023/apr/19/snapchat-making-ai-chatbot-similar-to-chatgpt-available-to-every-user. My AI now appears for Snapchat users as a contact in their social network, allowing users to ask it questions, have back and forth conversations, ask it to generate creative content, and much more. *Id.*

[98] Floyd, *supra* note 95; Silvia Pellegrino, *Which Companies Have Partnered With OpenAI*, TECHMONITOR (Jan. 18, 2023), https://techmonitor.ai/technology/which-companies-have-partnered-with-openai; Asif Iqbal, *OpenAi's Collaborations: Pushing the Boundaries of AI in Various Sectors*, LINKEDIN (Mar. 12, 2023), https://www.linkedin.com/pulse/openais-collaborations-pushing-boundaries-ai-various-sectors-iqbal/.

[99] Iqbal, *supra* note 98 ("In 2020, OpenAI announced a collaboration with drug manufacturer, Pfizer, to develop new AI technologies for drug discovery.").

[100] Danni Button, *ChatGPT Poses Danger for Online Dating Apps*, THE STREET (Feb. 15, 2023), https://www.thestreet.com/social-media/chatgpt-poses-dangers-for-online-dating-apps.

presents "the most complicated tech policy discussion possibly that [the country] has ever had."[101]

***"I am confident AI will be used by bad actors, and yes it will cause real damage."***[102] - Michael Schwarz, Microsoft's Chief Economist

***"If law and due process are absent from this field, we are essentially paving the way to a new feudal order of unaccountable reputational intermediaries."*** - Professors Danielle Keats Citron and Frank Pasquale at 2023 Geneva Conference.[103]

***AI technology is so powerful that it even has the potential to "allow an evil country, competitor to come in and screw up our democracy."***[104] - Eric Schmidt, Former Google CEO and Chairman at the 2023 Milken Global Conference.

93. In a report addressed to the American public in 2021, Eric Schmidt and Robert Work, the chair and vice chair of the National Security Commission on Artificial Intelligence ("NSCAI"), noted that "Americans have not yet grappled with just how profoundly the artificial intelligence revolution will impact our economy, national security, and welfare. Much remains to be learned about the power and limits of AI technologies. Nevertheless, ***big decisions need to be made now***…to defend against the malignant uses of AI."[105]

94. The NSCAI report highlights the consequences associated with the unregulated development of AI, emphasizing the unique risks to human rights, privacy, and personal autonomy. Further, the report notes the urgency of establishing comprehensive privacy frameworks and regulations that strike a balance between protecting individuals' privacy rights and enabling AI advancements.

95. On March 30, 2023, a new complaint was filed to the Federal Trade Commission

---

[101] Ben Wershkul & Alexandra Garfinkle, *White House bringing Google, Microsoft CEOs together for 'frank discussion' of AI*, YAHOO! FIN. (May 4, 2023), https://www.aol.com/finance/white-house-bringing-alphabet-microsoft-164428066.html.

[102] Bryce Baschuk, *Microsoft Economist Warns Bad Actors Will Use AI to Cause Damage*, MSN (May 3, 2023), https://www.msn.com/en-us/money/other/ai-will-cause-real-damage-microsoft-chief-economist-warns/ar-AA1aFslV.

[103] *EPIC AI Rulemaking Petition*, EPIC, https://epic.org/documents/epic-ai-rulemaking-petition/ (last visited February 14, 2024).

[104] Wershkul, *supra* note 101.

[105] Eric Schmidt & Bob Work, *Letter from the Chair and Vice Chair*, NAT'L. SEC. COMM'N. ON A.I., (2021), https://reports.nscai.gov/final-report/chair-and-vice-chair-letter.

("FTC"), urging the agency to investigate OpenAI and suspend its commercial deployment of large language models, including its latest iteration of the popular tool ChatGPT.[106] The complaint notes that the use of AI should be "transparent, explainable, fair, and empirically sound while fostering accountability."[107] None of the Products satisfy these requirements.

96. The significance of harm facing our society is in fact so imminent that Geoffrey Hinton—referenced by many as the "godfather" of AI—quit his job at Google where he had worked for more than a decade, becoming one of the most respected voices in the field, so he could freely speak out about the dangers associated with the rapid, uncontrolled development and release of AI to our society.

97. Dr. Hinton's journey from AI groundbreaker to AI whistleblower marks a remarkable moment for the AI technology industry at perhaps its most important inflection point in decades. Industry leaders believe the new AI systems could be as important but yet as catastrophic as the development of nuclear weapons.

98. After OpenAI released ChatGPT in March, more than 1,000 technology leaders and researchers signed an open letter calling for a six-month moratorium on the development of new systems because AI technologies pose "profound risks to society and humanity."[108]

99. Several days later, 19 current and former leaders of the Association for the Advancement of Artificial Intelligence, a 40-year-old academic society, released their own letter warning of the risks of AI. That group included Eric Horvitz, chief scientific officer at Microsoft, which has deployed OpenAI's technology across a wide range of products, including its Bing search engine.[109]

100. 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 . . .**

---

[106] Federal Trade Commission, *In the matter of OpenAI, Inc.*, FED. TRADE. COMM'N. (Mar. 30, 2023), https://cdn.arstechnica.net/wp-content/uploads/2023/03/CAIDP-FTC-Complaint-OpenAI-GPT-033023.pdf.

[107] *Id.*

[108] *The 'Godfather of A.I.' Leaves Google and Warns of Danger Ahead*, DNYUZ (May 1, 2023), https://dnyuz.com/2023/05/01/the-godfather-of-a-i-leaves-google-and-warns-of-danger-ahead/.

[109] *Id.*

> **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.[110]

101. 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; . . ."[111]

102. 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 ChatGPT. As OpenAI explained: novel capabilities often emerge in more powerful models. Some that are particularly concerning are the ability to create and act on long-term plans, to accrue power and resources ("power-seeking"), and to exhibit behavior that is increasingly "agentic."[112]

103. In February 2020, a petition with the Federal Trade Commission called on the FTC to conduct rulemaking for the use of artificial intelligence in commerce. "Given the scale of commercial AI use, the rapid pace of AI development, and the very real consequences of AI-enabled decision-making for consumers, [courts] should immediately initiate a rulemaking to define and prevent consumer harms resulting from AI."[113]

104. Multiple sources have called on the FTC to enforce the AI standards established in the OECD AI Principles, the OMB AI Guidance, and the Universal Guidelines for AI. Several FTC Commissioners have already acknowledged the FTC's role in regulating the use of AI.

105. The absence of effective AI regulations in the United States has accelerated the spread

---

[110] *Pause Giant AI Experiments: An Open Letter*, FUTURE OF LIFE INST. (Mar. 29, 2023), https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (emphasis in the original).
[111] *Id.*
[112] Dennis Layton, *GPT-4 – Some First Impressions*, LINKEDIN (Mar. 15, 2023), https://www.linkedin.com/pulse/gpt-4-some-first-impressions-dennis-layton ("Agentic in this context does not intend to humanize language models or refer to sentience but rather refers to systems characterized by the ability to, *e.g.*, accomplish goals which may not have been concretely specified and which have not appeared in training; focus on achieving specific, quantifiable objectives; and [engage in] long-term planning.").
[113] *EPIC AI Rulmaking Petition*, *supra* note 103.

of unaccountable and untrustworthy AI tools. And the unregulated use of those AI tools has already caused serious harm to consumers, who are increasingly subject to opaque and unprovable decision-making in employment, credit, healthcare, housing, and criminal justice.

106. Realizing the gravity of potential harm, authorities within European countries took ChatGPT offline in Italy in April after the country's data protection authority temporarily banned the chatbot and launched a probe into the artificial intelligence application's suspected breach of privacy rules.[114]

107. Italian authorities stated that ChatGPT has an "absence of any legal basis that justifies the massive collection and storage of personal data" to "train" the chatbot.[115] Further, they accused Defendant OpenAI of failing to check the age of ChatGPT's users to ensure they are aged 13 or above.[116]

108. Subsequently, Defendant OpenAI agreed to offer specific tools to verify Users' ages in Italy upon sign-up, but yet continues to enable unverified access in the United States to illegally collect the personal data of minors. Defendant OpenAI also said that it would provide greater visibility of its privacy policy and user content opt-out form, creating a new form for European Union users to exercise their right to object to its use of personal data to train its models. The form requires people who want to opt out to provide detailed personal information, including evidence of data processing via relevant prompts. However, despite consumers' established privacy rights to be "forgotten," Defendants cannot effectively extract individuals' information from the Products once the AI is trained on such information.[117]

109. Italy was the first western European country to curb ChatGPT, but its rapid

---

[114] Supantha Mukherjee & Giselda Vagnoni, *Italy Restores ChatGPT After OpenAI Responds to Regulator*, YAHOO! (Apr. 28, 2023), https://finance.yahoo.com/news/chatgpt-available-again-users-italy-163139143.html.

[115] Elvira Pollina & Supantha Mukherjee, *Italy Curbs ChatGPT, Starts Probe Over Privacy Concerns*, REUTERS (Mar. 31, 2023), https://www.reuters.com/technology/italy-data-protection-agency-opens-chatgpt-probe-privacy-concerns-2023-03-31/.

[116] *Id.*

[117] *ChatGPT and Education*, CNT. FOR INNOVATIVE TEACHING AND LEARNING, https://www.niu.edu/citl/resources/guides/chatgpt-and-education.shtml, (last visited February 14, 2024) ("the prompts that you input into ChatGPT cannot be deleted. If you, or your students, were to ask ChatGPT about sensitive or controversial topics, this data cannot be removed.").

development has attracted attention from lawmakers and regulators in several countries. A committee of European Union lawmakers agreed on new rules that would force companies deploying generative AI tools, such as ChatGPT, to disclose any copyrighted material used to develop their systems.[118]

110. Data authorities from around the world remain concerned, specifically, with "the lack of legal basis underpinning the massive collection, use and disclosure of personal information in order to train the ChatGPT algorithms on which the platform relies" and the "cornerstone privacy issue" at the heart of this Action: ChatGPT's "use of web scraping and the collection of personal information without consent."[119]

111. In short, the message is consistent from informed business, nonprofit, and technology thought leaders; industrialists; scientists; world leaders; regulators; and governments around the globe: The proliferation of AI—including Defendants' products—pose an existential threat if not constrained by the reasonable guardrails of our laws and societal mores. Defendants' business and scraping practices raise fundamentally important legal and ethical questions that must also be addressed. Enforcing the law will not amount to stifling AI innovation, but rather a safe and just AI future for all.

**B. Overview of Risks**

112. The following is a brief, non-exhaustive list of ongoing harms and critical legal threats

---

[118] Supantha Mukherjee & Giselda Vagnoni, *Italy Restores CHATGPT after OpenAI Responds to Regulator*, SRN NEWS (Apr. 28, 2023), srnnews.com/italy-restores-chatgpt-after-openai-responds-to-regulator-2/.

[119] Roland Hung, *AI Technology and Privacy: Canadian Privacy Commissioner Launches Investigation into ChatGPT*, TORKIN MANES (Apr. 24, 2023), https://www.torkinmanes.com/our-resources/publications-presentations/publication/ai-technology-and-privacy-canadian-privacy-commissioner-launches-investigation-into-chatgpt (detailing the "privacy concerns with the use of ChatGPT" that have been raised worldwide). *See also* Heinrich Long, *Authorities Press OpenAI to Disclose How ChatGPT Input Is Used*, RESTORE PRIV. (June 9, 2023), https://restoreprivacy.com/authorities-press-openai-to-disclose-how-chatgpt-input-is-used/ (discussing worldwide investigations, including the latest inquiry from Dutch data protection authorities who "want[] to know, among other things, how OpenAI handles personal data when training the underlying system. The[y…] want[] to know from OpenAI whether people's questions are used to train the algorithm, and if so, in what way. The[y…] also ha[ve]questions about the way in which OpenAI collects and uses personal data from the internet.").

the Products pose to everyday Americans, including Plaintiff and the Proposed Class Members.

***1. Massive Privacy Violations***

113. In today's vast, interconnected digital landscape, privacy can appear to be more of an illusion, but it is still a guaranteed right. In violation of this right, the Products operate as an all-seeing online platform, tracking our every move: each click, each site visit, each chat—not allowing anything to escape its relentless scrutiny. Internet users' interactions, seemingly innocuous, are aggregated, filtered, and compiled by Defendants, rendering the concept of privacy virtually non-existent. Even information deemed private or intended for a restricted audience does not escape surveillance.

114. The massive, unparalleled collection and tracking of users' personal information by Defendants endangers individuals' privacy and security to an incalculable degree. This information can be exploited and used to perpetrate identity theft, financial fraud, extortion, and other malicious purposes. It can also be employed to target vulnerable individuals with predatory advertising, algorithmic discrimination, and other unethical and harmful acts.

115. The collection and use of this data raises concerns about user privacy and the potential misuse of personal information. For example, every iota of Users' activity is tracked and monitored. By analyzing this data using algorithms and machine learning techniques, Defendants can develop a chillingly detailed understanding of users' behavior patterns, preferences, and interests—creating an entirely new meaning to the term "invasive."

116. Several studies confirm that the collection and disclosure of sensitive information from millions of individuals, as Defendants have done here, violates established expectations of privacy based on long-standing social norms. Privacy polls and studies uniformly show that the overwhelming majority of Americans consider one of the most important privacy rights to be the need for an individual's affirmative consent before a company collects and shares its customers' data.

117. For example, a recent study by Consumer Reports reveals that 92% of Americans believe that internet companies and websites should be required to obtain consent before selling or sharing consumers' data, and that internet companies and websites should be required to provide

consumers with a complete list of the data that has been collected about them.[120] Moreover, according to a study by Pew Research Center, a majority of Americans, approximately 79%, are concerned about how companies collect data about them.[121]

118. Users act consistently with these privacy preferences. Following a new rollout of the iPhone operating software—which asks users for clear, affirmative consent before allowing companies to track users—85% of worldwide users and 94% of U.S. users chose not to share data when prompted.[122] The Products' Users do not have that option, and do not understand the full extent of Defendants' data collection and use of their personal data.

119. While the reams of personal information that Defendants collect on Users can be used to provide personalized and targeted responses, it can also be used for exceedingly nefarious purposes, such as tracking, surveillance, and crime. For example, if ChatGPT has access to a User's browsing history, search queries, and geolocation, and combines this information with what Defendant OpenAI has secretly scraped from the internet, Defendants could build a detailed profile of Users' behavior patterns, including but not limited to where they go, what they do, with whom they interact, and what their interests and habits are. This level of surveillance and monitoring raises vital ethical and legal questions about privacy, consent, and the use of personal data. It is crucial for users to be aware of how their data is being collected and used, and to have control over how their information is shared and used by advertisers and other entities.

120. The concern about collecting and sharing information is compounded by the reality that this information may include particularly sensitive information such as medical records or information about minors. Increasingly, companies like Defendants "are harnessing and collecting multiple typologies of children's data and have the potential to store a plurality of data traces under

---

[120] *Consumers Less Confident About Healthcare, Data Privacy, and Car Safety, New Survey Finds*, CONSUMER REPS. (May 11, 2017), https://www.consumerreports.org/consumer-reports/consumers-less-confident-about-healthcare-data-privacy-and-car-safety/.

[121] Brooke Auxier et al., *Americans and Privacy: Concerned, Confused, and Feeling Lack of Control over Their Personal Information*, PEW RSCH. CTR. (Nov. 15, 2019), https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

[122] Margaret Taylor, *How Apple Screwed Facebook*, WIRED (May 19, 2021, 6:00 AM), https://www.wired.co.uk/article/apple-ios14-facebook.

unique ID profiles."[123]

121. Given ChatGPT's ability to generate human-like understanding and responses, there is a high likelihood that users might share (and already are sharing) their private health information while interacting with the model, by asking health-related questions or discussing their medical history, symptoms, or conditions. Moreover, this information can be logged and reviewed as part of ongoing efforts to "train," improve and monitor each model's performance.

122. However, beyond these seemingly innocuous interactions with the AI, healthcare industry providers are beginning to integrate ChatGPT in order to "revolutionize healthcare" while undermining the confidentiality of individuals' personal data, which would be transmitted using ChatGPT and continuing to train Defendants' AI at the patients' expense.[124] While this technology could provide benefits, the risks associated with its implementation are drastic, from cybercrime, misinformation and misdiagnosis, lack of empathy and experience, and bias[125] to the existential risk, of which Altman has repeatedly warned.

123. ***Established Privacy Rights to be "Forgotten" Violated.*** Compounding this massive invasion of privacy, OpenAI offers no *effective* procedures at this time for individuals to request for their information/training data to be deleted. Instead, OpenAI simply provides an email address that consumers can contact if they would like to have their information removed. But this "option" is illusory. Regardless of whether individuals can technically request for ChatGPT to remove their data, it is not possible to do so completely, because Defendants train ChatGPT on individual inputs, personal information, and other user and nonuser data, which Defendants cannot reliably and fully extract from its trained AI systems any more than a person can "unlearn" the math they learned in sixth grade.

124. An AI researcher with privacy and cybersecurity firm AVG explains, "People are

---

[123] Veronica Barassi, *Tech Companies Are Profiling Us from Before Birth*, THE MIT PRESS READER, (Jan. 14, 2021), https://thereader.mitpress.mit.edu/tech-companies-are-profiling-us-from-before-birth/.

[124] Naomi Diaz, *6 Hospitals, Health Systems Checking Out ChatGPT*, BECKERS HEALTHCARE (June 2, 2023), https://www.beckershospitalreview.com/innovation/4-hospitals-health-systems-testing-out-chatgpt.html.

[125] Ethan Popowitz, *ChatGPT: Friend or Foe?*, DEFINITIVE HEALTHCARE, https://www.definitivehc.com/blog/chatgpt (last visited February 14, 2024).

furious that data is being used without their permission. . . Sometimes, some people have deleted the[ir] [online] data but since the language model has already used them, the data is there forever. They don't know how to delete the data."[126]

125. Likewise, some companies have banned or limited ChatGPT use because they are "worried that anything uploaded to AI platforms like OpenAI's ChatGPT or Google's Bard will [also] get *stored* on those companies' servers, *with no way to access or delete the information.*"[127]

126. The "right to be forgotten"—*i.e.*, the right to request that a business delete the personal information that it holds about you—is guaranteed to California residents under the California Consumer Privacy Act of 2018 ("CCPA"). Given how the technology works, OpenAI is not compliant with these requirements.[128]

### *2. AI-Fueled Misinformation Campaigns, Targeted Attacks, Sex Crimes, and Bias*

127. ***Misinformation, Deepfakes, Clones, Scams, and Blackmail*:** The use of the Products facilitates the spreading of false or misleading information, even without "misuse." That is because a *feature* (known defect) of ChatGPT's *regular use* is the inventing of false information, including potentially defamatory information about individuals. Even the "improved" version (GPT4) "makes stuff up" and "may generated text that is completely false."[129]

128. One high-profile example involves a US law professor, Jonathan Turley, who ChatGPT falsely accused of sexually harassing one of his students, even providing a "source" for

---

[126] *Is ChatGPT's use of people's data even legal?*, AVG, https://www.avg.com/en/signal/chatgpt-data-use-legal? (last visited February 14, 2024).

[127] Felicity Nelson, *Many Companies are Banning ChatGPT. This is Why*, SCI. ALERT (June 16, 2023), https://www.sciencealert.com/many-companies-are-banning-chatgpt-this-is-why (emphasis added). Microsoft has itself directed employees not to share sensitive data with ChatGPT "in case it's used for future AI training models" Diamond Naga Siu, *Microsoft is chill with employees using ChatGPT — just don't share 'sensitive data' with it*, YAHOO! NEWS (Feb. 1, 2023), https://news.yahoo.com/microsoft-chill-employees-using-chatgpt-114000174.html?guccounter=1.

[128] *See, e.g.,* Alexa Johnson-Gomez, *A "Living" AI: How ChatGPT Raises Novel Data Privacy Issues*, MINN. J. OF L., SCI. & TECH. BLOG (Feb. 6, 2023), https://mjlst.lib.umn.edu/2023/02/06/a-living-ai-how-chatgpt-raises-novel-data-privacy-issues/ (dismissing purported compliance with CCPA as "in name only" given how the data is used as part of machine learning model).

[129] Cade Metz, *10 Ways GPT-4 is Impressive but Still Flawed*, THE N.Y. TIMES (Mar. 14, 2023), https://www.nytimes.com/2023/03/14/technology/openai-new-gpt4.html.

the purported crime via a news article that it invented.[130] Defendants call this "hallucination," but the world knows it as defamation. While Defendants are allegedly "working on" a fix for this behavior, they continue to push the defective Product worldwide. Naturally, one would expect an ethical company "for the *benefit* of humanity" not to release such a Product, at all, *unless and until* it was safeguarded from committing crimes *against* humanity.

129. The Cambridge Analytica scandal—in which personal data was allegedly misused to target individuals with political propaganda and misinformation—is also an instructive cautionary tale.[131] Cambridge Analytica collected personal data using third-party apps that collected data from users and their friends. It then used this data to build detailed profiles of individuals, so they could be targeted with personalized political ads and propaganda. Cambridge Analytica used algorithms and machine learning techniques to analyze this data, identify patterns in users' behavior and preferences, and target those users with specific messages and ads.

130. This history highlights the potential dangers of using personal data to build detailed profiles of individuals, particularly when that data is collected without their knowledge or consent. It also raises important questions about the ethics of using personal data for political purposes and the need for greater regulation and oversight of data collection and use.

131. Moreover, by allowing the collection, storage, and analysis of a massive amount of highly individualized, personal data—from audio and photographic data to detailed interests, habits, and preferences—OpenAI's technology facilitates the proliferation of video or audio "deepfakes" and makes them harder to detect.[132] Simply put, the Products make it easier to create lifelike audiovisual digital duplicates—digital clones—of real people, which can then be used to spread

---

[130] Hern and Milmo, *supra* note 59.

[131] Sam Meredith, *Here's Everything You Need to Know About the Cambridge Analytica Scandal*, CNBC (Mar. 21, 2018), https://www.cnbc.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html. (The Cambridge Analytica scandal involved the misuse of personal data collected from Facebook users, which was then used to target individuals with political advertising and propaganda. The scandal highlighted the potential dangers of using personal data for targeted advertising and the need for greater transparency and accountability in the collection and use of personal information.).

[132] Bibhu Dash & Pawankumar Sharma, *Are ChatGPT and Deepfake Algorithms Endangering the Cybersecurity Industry? A Review*, 10(1) I. J. OF ENG'G & APPLIED SCI. (Jan. 2023), https://www.ijeas.org/download_data/IJEAS1001001.pdf.

misinformation, exploit victims, or even access privileged data.[133]

132. Deepfakes could influence elections, erode public trust, and negatively affect public discourse.[134] The U.S. Congressional Research Service has further analyzed the risks of deepfakes, explaining that they could be used to "blackmail elected officials or individuals with access to classified information" and "generate inflammatory content […] intended to radicalize populations, recruit terrorists, or incite violence.[135]

133. In addition to spreading misinformation, criminals have used, and will continue to use this technology to harass, blackmail, extort, coerce, and defraud. Armed with artificial intelligence tools like the ones developed by Defendants, malicious actors can weaponize even the most innocuous publicly available personal information, such as names and photographs, against private individuals.

134. For example, the FBI has issued an alert about a particularly despicable form of blackmail currently on the rise that has been largely facilitated by AI like the Products. This scheme, a form of "sextortion," is perpetrated using artificial intelligence tools and publicly available photographs and videos of private individuals, usually obtained through social media, to create deepfakes containing pornographic content.[136] The photos or videos are then publicly circulated on social media, public forums, and pornographic websites for the purpose of harassing the victim, causing extreme emotional and psychological distress.[137]

135. A malicious actor may also attempt to extract ransom payments, sometimes seeking genuine versions of the subject engaging in the acts depicted in the made up sexually-explicit images and videos, by threatening to share the falsified images or videos with family members, social contacts, or by indiscriminately circulating the content on social media.[138] The most concerning and

---

[133] *Science & Tech Spotlight: Deepfakes*, U.S. GOV'T ACCOUNTABILITY OFF. (Feb. 20, 2020), https://www.gao.gov/products/gao-20-379sp; *see also* Dash & Sharma, *supra* note 132.
[134] Kelley M. Sayler & Laurie A. Harris, *Deep Fakes and National Security*, CONG. RSCH. SERV., (April 17, 2023), https://crsreports.congress.gov/product/pdf/if/if11333.
[135] *Id.*
[136] *Public Service Announcement: Malicious Actors Manipulating Photos and Videos to Create Explicit Content and Sextortion Schemes*, FED. BUREAU OF INVESTIGATION (June 5, 2023), https://www.ic3.gov/Media/Y2023/PSA230605.
[137] *Id.*
[138] *Id.*

egregious aspect of this type of "sextortion" scheme is that the victims include not only non-consenting adults, but also minor children.[139]

136. ***Child Pornography.*** Defendants' Product Dall-E has become a favorite tool for pedophiles, because it requires less technical competence than previous programs used by pedophiles and increases the scale at which images of virtual child pornography can be created.[140] In just mere seconds, Dall-E can create realistic images of children performing sex acts.[141] Thousands of such images have already been detected in dark web forums.[142] In a dark web forum with 3,000 subscribers, 80% of respondents to an internal poll stated "they had used or intended to use AI tools to create child sexual abuse images."[143] In such forums, users exchange strategies for thwarting the woefully insufficient purported "safety guardrails" of Dall-E and other AI products, "including by using non-English languages they believe are less vulnerable to suppression or detection."[144]

137. Dall-E is a diffusion model, and anyone can access it, generating a realistic image solely by typing a short description of the desired product. [145] This model was trained off billions of images taken, without notice or consent, from the internet, "many of which showed real children and came from photo sites and personal blogs."[146] Images of actual children are thus the source material for the AI-generated child pornography. In some instances, actual images of existing child pornography were used to train the model and generate further explicit material of already victimized children, thereby victimizing them all over again.[147]

138. AI-generated child pornography has introduced a slew of other horrendous problems as well. "The flood of images could confound the central tracking system built to block such material

---

[139] *Id.*
[140] Drew Harwell, *AI-generated Child Sex Images Spawn New Nightmare for the Web*, WASH. POST (June 19, 2023, 7:00 AM), https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images/.
[141] *Id.*
[142] *Id.*
[143] *Id.*
[144] *Id.*
[145] *Id.*
[146] *Id.*
[147] *Id.*

from the web because it is designed only to catch known images of abuse, not detect newly generated ones."[148] Moreover, the monumental task of locating children harmed by the production of child pornography has been bogged down now that agents must now spend time puzzling over whether content is real or virtual.[149] Furthermore, this virtual material is not merely used by pedophiles to supplant real material.[150] AI is also being used to "build [] fake school-age persona[s]" via fabricated selfies, which are incorporated into plots to lure and groom child targets.[151]

139. Absent the injunctive relief sought in this action, Defendants will continue to not only steal data from unwitting victims, including minors, but arm pedophiles in rapidly generating child pornography at scale and in creating materials that can be strategically used to groom and victimize real children.

140. ***Hate and Bias.*** Continued commercial deployment of the Products also will amplify and entrench the human biases and prejudices reflected in the Products' sources, which Defendants used without regard to such factors by incorporating and training the Products with content from various extremist websites and by failing to use adequate filtering safeguards.[152]

### *3. Hypercharged Malware Creation*

141. ***Malicious, Mutating, and Virtually Undetectable Code Scripts*:** Malware, or malicious software, are computer programs designed to damage or infiltrate computer systems. Unscrupulous actors deploy malware by embedding them within vulnerabilities in existing internet applications.[153] The Products guarantee that "malware" prevalence and potency will exponentially

---

[148] *Id.*
[149] *Id.*
[150] *Id.*
[151] *Id.*
[152] Sam Biddle, *The Internet's New Favorite AI Proposes Torturing Iranians and Surveilling Mosques*, THE INTERCEPT (Dec. 8, 2022), https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/.
[153] Fei Xiao et al., *A Novel Malware Classification Method Based on Crucial Behavior*, 2020 MATHEMATICAL PROBS. IN ENG'G. (Mar. 21, 2020), https://doi.org/10.1155/2020/6804290; Rabia Tahir, *A Study on Malware and Malware Detection Techniques*, 2 INT'L J. OF MGMT. ENG'G., 20, 20 (Mar. 8, 2018), https://www.mecs-press.net/ijeme/ijeme-v8-n2/IJEME-V8-N2-3.pdf; Mohd Faizal Ab Razak et al., *The Rise of "Malware": Bibliometric Analysis of Malware Study*, 75 J. OF NETWORK AND COMPUT. APPLICATIONS, 58, 58 (Nov. 2016), https://www.sciencedirect.com/science/article/pii/S1084804516301904.

increase, posing unprecedented cybersecurity risks on a global scale. That is because the Products can generate virtually undetectable malware, and at massive scale, to thwart security systems and jeopardize entire governments.

142. Malware attacks have sabotaged entire governments before. For example, in 2022, the Russian Conti Group enacted a weeks-long attack on 27 different ministries in the Costa Rican government.[154] The malware deployed was ransomware, a software that encrypts critical information, denying access to its rightful owner and threatening its destruction if payment is not made.[155] Costa Rica's president declined to pay the $20M ransom, but a standoff ensued leaving parts of Costa Rica's digital infrastructure in shambles, disrupting public healthcare and the pay of its workers.[156]

143. Healthcare providers are also often targeted by malware, and increasingly so. For example, a major software provider for the UK's National Health System sustained a ransomware attack from an unknown group last summer.[157] The attack had real impact on the health of millions, disrupting ambulance dispatch, appointment scheduling, and emergency prescriptions, among other things.[158] Ransomware attacks on health care providers have doubled from 2016 to 2021, exposing the sensitive health information of 42M individuals.[159]

144. ***The Products supercharge Malware:*** In 2012, 33% of malware went undetected by antivirus software.[160] In the last decade, malware has become ever more sophisticated, and ever more capable of thwarting detection. But now, with the assistance of the Products, malware can

---

[154] Christine Murray & Mehul Srivastava, *How Conti Ransomware Group Crippled Costa Rica-Then Fell Apart*, FIN. TIMES (July 9, 2022), https://www.ft.com/content/9895f997-5941-445c-9572-9cef66d130f5.

[155] *Id.*

[156] *Id.*

[157] Vedere Labs, *Ransomware in Healthcare: The NHS Example and What the Future Holds*, SEC. BOULEVARD (Aug. 25, 2022), https://securityboulevard.com/2022/08/ransomware-in-healthcare-the-nhs-example-and-what-the-future-holds/.

[158] *Id.*

[159] Hannah T. Neprash et al., *Trends in Ransomware Attacks on US Hospitals, Clinics, and Other Health Care Delivery Organizations, 2016–2021,* JAMA HEALTH FORUM (Dec. 29, 2022), https://jamanetwork.com/journals/jama-health-forum/fullarticle/2799961.

[160] Markus Kammerstetter et al., *Vanity, Cracks, and Malware: Insights into the Anti-Copy Protection Ecosystem*, ASS'N. FOR COMPUTING MACHINERY 809, 818 (Oct. 16, 2012), https://doi.org/10.1145/2382196.2382282.

become undetectable in new ways, at scale, because ChatGPT can be used to create "mutating, or polymorphic" malware.[161] Polymorphic malware has a mutation engine with self-propagating code that allows it to rapidly change its appearance and composition.[162] This malware can change its entire make-up, so that malware detectors, reactionary by nature, will not recognize its newer, ongoing permutations.[163]

145. ChatGPT can build the requisite polymorphic code, using its API at runtime to deploy advanced malware attacks that evade detection by security systems designed to thwart malware, such as endpoint detection and response (EDR) applications.[164] Recently, researchers designed a simple, executable file that corresponds with ChatGPT's API in real time "to generate dynamic, mutating versions of malicious code," making it extremely difficult to detect using existing cybersecurity tools.[165]

146. While the most recent iterations of ChatGPT purport to "disallow" potential prompt injections for generating polymorphic malware, this supposed guardrail for safety is woefully inadequate: cleverly worded inputs, used by developers of malware, easily circumvent ChatGPT's content filters with a practice commonly referred to as "prompt engineering."[166]

147. Thus, Mackenzie Jackson, developer advocate at cybersecurity company GitGuardian warns that, as generative models become more advanced, "AI may end up creating malware that can only be detected by other AI systems for defense. What side will win at this game is anyone's guess."[167] To knowingly put this enhanced ability to sabotage governments, health care systems, and any other number of targets into the hands of everyday people worldwide without adequate safeguards is emblematic of Defendants' gross negligence and underscores the need for immediate judicial intervention.

---

[161] Shweta Sharma, *ChatGPT Creates Mutating Malware That Evades Detection by EDR*, CSO ONLINE (June 6, 2023, 1:59 PM), https://www.csoonline.com/article/3698516/chatgpt-creates-mutating-malware-that-evades-detection-by-edr.html.
[162] *Id.*
[163] *Id.*
[164] *Id.*
[165] *Id.*
[166] *Id.*
[167] *Id.*

### *4. Autonomous Weapons*

148. AI also poses a unique threat to international security and human rights through the development of autonomous weapons known as "Slaughterbots," otherwise known as "lethal autonomous weapons systems" or "killer robots," which are weapons systems that use AI to identify, select, and kill human targets without intervention.[168] As one humanitarian organization explained, "[w]eapons that use algorithms to kill, rather than human judgment, are immoral and a grave threat to national and global security."[169]

149. The risk that unregulated AI like the Products pose via autonomous weapons is "not a far-fetched danger for the future, but a clear and present danger."[170] Such weapons have already nearly killed a foreign head of state, and due to the rapid commercial proliferation of open-source AI, "could be built today by an experienced hobbyist for less than $1,000."[171]

150. Defendants' conduct exacerbates the problem. There is already an early autonomous implementation of ChatGPT known as "Chaos GPT" which is being touted as "empowering GPT with Internet and Memory to Destroy Humanity."[172] Chaos-GPT is predicated on an open source application that uses Defendants' GPT-4, and was designed by an anonymous user to be a "destructive, power-hungry, manipulative AI."[173] With only those parameters set by the user, Chaos-GPT returned a list of objectives it set for itself. One was to "destroy humanity." Another was to "cause chaos and destruction" by creating "widespread suffering."[174] Next, Chaos-GPT, of its own "volition," prepared a plan in support of these objectives – and then it searched the internet for weapons of mass destruction seeking to obtain one.[175]

---

[168] *See Slaughterbots Are Here*, AUTONOMOUS WEAPONS (Feb. 23, 2023), https://autonomousweapons.org/ (discussing Latin American and the Caribbean Conference on the Social and Humanitarian Impact of Autonomous Weapons).
[169] *Id.*
[170] Kai-Fu Lee, *The Third Revolution in Warfare*, THE ATLANTIC (Sept. 11, 2021), https://www.theatlantic.com/technology/archive/2021/09/i-weapons-are-third-revolution-warfare/620013/.
[171] *Id.*
[172] Jose Antonio Lanz, *Meet ChaosGPT: An AI Tool That Seeks to Destroy Humanity*, DECRYPT (Apr. 13, 2023), https://decrypt.co/126122/meet-chaos-gpt-ai-tool-destroy-humanity.
[173] *Id.*
[174] *Id.*
[175] *Id.*

151. Experts warn that advancements in AI like those accomplished by the Products, "will accelerate the near-term future of autonomous weapons."[176] While it is believed artificial intelligence at a level equal to or higher than human intelligence is a prerequisite to truly autonomous weaponry, the unfettered commercial deployment of the Products naturally escalates this risk as their widespread use continually "enhances" the AI's capabilities – and without sufficient moral or ethical guardrails, as sought in this Action.

**C. Opportunity on the Other Side**

152. While leading experts agree on the grave risks posed by the Products, and the need for a temporary pause in their commercial deployment, it is important to understand the full picture of why this Action matters. It is not just to contain the risks to society and harms happening right now, including the supercharged spread of disinformation, the obliteration between truth and fiction, deepfakes designed to harass, harm, and commit fraud, and more. It is not just to halt Defendants' ongoing disregard for the privacy and property interests of millions, and to remedy those violations. It is not just to avoid the collapse of civilization as we know it and as Mr. Altman himself recognizes is possible.[177] Naturally, all of these things warrant the comparatively measured relief Plaintiff and the Classes seek. But beyond all of this, the Action matters to ensure humankind can *realize the tremendous opportunity for advancement and prosperity* that awaits us, on the other side of a commercial pause.

153. By pausing now, "[h]umanity can enjoy a flourishing future."[178] It will enable the joint development and implementation of shared safety protocols, overseen by independent outside experts, to manage the risks and render the Products safe to usher in an exciting new era of progress for all. For example, with adequate safeguards, the Products will be positioned to revolutionize healthcare for good, by helping to discover new drugs to save lives and potentially find cures for

[176] Lee, *supra* note 170.

[177] David Meyer, *Sam Altman Has Signed a New Open Letter on A.I.'s Dangers: Here's What's Different About This 'Extinction' Statement*, FORTUNE MAG. (May 30, 2023, 9:55 AM), https://fortune.com/2023/05/30/sam-altman-has-signed-a-new-open-letter-on-a-i-s-dangers-heres-whats-different-about-this-extinction-statement/.

[178] *Pause Giant AI Experiments: An Open Letter*, FUTURE OF LIFE INST. (Mar. 22, 2023), https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

cancer and other deadly diseases. With adequate safeguards, the Products can contribute not only to our everyday efficiency, artistic expression, joy and more, but also to the greater societal good by advancing human rights, promoting social justice, reducing inequities, and empowering marginalized groups.

154. With adequate safeguards, including a moral and ethical code, the Products can help detect and prevent human rights violations rather than cause them; they can help combat human discrimination and bias rather than replicate, encourage, and exacerbate humankind's worst impulses.[179] On the other side of the pause, the Products can responsibly foster global cooperation, collaboration, and peace by facilitating communication, learning, and understanding across cultures and languages rather than starting world wars with disinformation and the unchecked capacity for autonomous weaponry. Likewise, the Products can aid in the ongoing search for truth, by enabling breakthroughs in math, science, and more, that humans might never alone make, rather than forever obliterating the line between truth and fiction altogether.

155. We can have this AI, the one that enriches our lives, that works for people, and that works for human benefit, that is "helping us cure cancer, that is helping us find climate solutions," but leading experts agree, not without a pause on the Products' unchecked commercial proliferation: "[W]hen we're in an arms race to deploy AI to every human being on the planet as fast as possible with as little testing as possible, that's not an equation that's going to end well."[180] The current scenario stands only to enrich Defendants, while destabilizing the world.

156. Defendants have released Products to the entire world, that they know and readily recognize could someday result in societal collapse; that even they, the creators, cannot fully understand, predict, or reliably control; thus, any attempt now by Defendants to politicize this

[179] *See generally* Cade Metz and Karen Weise, *A Tech Race Begins as Microsoft Adds A.I. to Its Search Engine*, THE N.Y. TIMES (Feb. 7, 2023), https://www.nytimes.com/2023/02/07/technology/microsoft-ai-chatgpt-bing.html ("The new chatbots do come with baggage. They often do not distinguish between fact and fiction. They can generate language that is biased against women and people of color. And experts worry that people will use them to spread lies at a speed they could not in the past.").

[180] Jason Abbruzzese, *The Tech Watchdog That Raised Alarms About Social Media Is Warning About AI*, NBC NEWS (Mar. 22, 2023), https://www.nbcnews.com/tech/tech-news/tech-watchdog-raised-alarms-social-media-warning-ai-rcna76167.

action, to attack the class action device or those brave enough to stand up to corporate greed and irresponsibility of this magnitude at this pivotal moment in history, will fail. All people of good will on both sides of the aisle and from every background are united and resolute in the need for intervention. That is because we all want to live in a world where technology serves our shared values of freedom, justice, dignity, equality, prosperity, privacy and security, not where Products exist that undermine these ideals.

157. In an often divided and polarized world, it is telling how so many have been able to unite around these truths: (i) the current state of AI governance is insufficient to address the threats posed by the Products; (ii) the lack of transparency, accountability, oversight, and regulation surrounding the Products and Defendants suddenly deploying them for profit worldwide has resulted in a ticking time bomb in the hands of those motivated to harm the American people; (iii) the gap must be closed between the rapid pace of the Products' development on the backs of stolen personal data on the one hand, and the slow progress of AI policy on the other; and (iv) a temporary pause on the commercial deployment of the Products is necessary and justified to prevent irreversible damage to humanity and society.

158. Critically, the injunctive relief sought in this Action seeks only to pause the unfettered and further commercial deployment of the Products, with AI research and development otherwise continuing unaffected. That is because of an equally important truth on which all agree: the United States must remain aggressively locked into the worldwide AI arms-race, set off by Defendants' launch of the Products (for better or worse), to ensure this powerful technology is developed and deployed for good around the world, and to block the potential harms from those world powers currently leveraging AI like the Products to build technological weapons as powerful as the nuclear bomb. Thus, the only "setback" here will be to Defendants' corporate bank accounts, while the American people stand to (re)gain their fundamental right to privacy as well as just compensation for the mass theft of personal data on which Defendants built and continue to run the Products.

## III. DEFENDANTS' CONDUCT VIOLATES ESTABLISHED PROPERTY AND PRIVACY RIGHTS

### A. Defendants' Web-Scraping Theft

159. Defendants' first category of theft and misappropriation stems from their secret scraping of the internet. This violated both the property rights and privacy rights of all individuals whose personal information was scraped and then incorporated through misappropriation into Defendants' Products.

160. Defendants' initial web scraping was done largely in secret, without the consent of any individuals whose personal and identifying information was scraped, much less all of the website operators themselves. This violated not only the Terms of Use of various websites but also the rights of each and every individual to opt out of such collection under California and other state and federal laws. Without any notice to the public, no one can be said to have consented to the collection of their online personal data, history, web practices and other personal and identifying information.

161. By the time the public learned of Defendants' web scraping practices in late Fall of 2022, when ChatGPT was released, it was too late to meaningfully exercise their privacy rights outside of this lawsuit — their internet history had been scraped, consumed, and integrated into the large language models from which the Products were born.

162. While Defendants' massive theft of personal information at scale is unmatched in history, it is reminiscent of the Clearview AI scandal in 2020. Clearview is a company that uses facial recognition technology to identify individuals based on their online photos.[181] To create its product, Clearview scraped billions of publicly available photos from various websites and social media platforms.[182] As with Defendants, this illegal scraping was done without the consent of users or the website owners themselves, and without registering as a data broker under California or

[181] Tate Ryan-Mosley, *The NYPD Used a Controversial Facial Recognition Tool. Here's What You Need to Know*. MIT TECH. REV., (Apr. 9, 2021), www.technologyreview.com/2021/04/09/1022240/clearview-ai-nypd-emails/.
[182] Will Knight, *Clearview AI Has New Tools to Identify You in Photos*, WIRED (Oct. 4, 2021), https://www.wired.com/story/clearview-ai-new-tools-identify-you-photos/.

Vermont Law.[183]

163. Just like Defendants, Clearview used the stolen information to build its AI product.[184] Clearview then sold access to the product to law enforcement agencies, private companies, and other governmental agencies.[185] Defendants' business model is the same: scrape information off the internet, in secret without any notice and consent in violation of the law, use it to build AI products, and then sell access to the Products for commercial gain.

164. Clearview's illegal scraping practices also went undetected for years, until it was laid bare by a New York Times expose.[186] The public was rightfully upset, as were state and federal regulators. The Vermont Attorney General sued Clearview in March 2020 for violating data broker and consumer protection laws, alleging that Clearview fraudulently acquired brokered personal information through its scraping practices and exposed consumers to various risks and harms.[187] Clearview was also sued by several individuals and organizations in California and elsewhere.[188]

165. As a result of these lawsuits and public scrutiny, Clearview ultimately registered as a data broker in both California and Vermont. Although Defendants employ the same business model as Clearview, they have failed to register as data brokers under applicable law. By failing to do so prior to scraping the internet, Defendants violated the rights of millions. Plaintiff and the Classes had a right to know what personal information Defendants were scraping and collecting and how it would be used, a right to delete their personal information collected by Defendants, and a right to

---

[183] Robert Hart, *Clearview AI Fined $9.4 Million in UK for Illegal Facial Recognition Database*, FORBES (May 23, 2022), https://www.forbes.com/sites/roberthart/2022/05/23/clearview-ai-fined-94-million-in-uk-for-illegal-facial-recognition-database/?sh=73d5a0f71963.

[184] *Id.*

[185] Drew Harwell, *Clearview AI to Stop Selling Facial Recognition Tool to Private Firms*, THE WASH. POST (May 9, 2022), https://www.washingtonpost.com/technology/2022/05/09/clearview-illinois-court-settlement/.

[186] Dave Gershgorn, *Is There Any Way Out of Clearview's Facial Recognition Database?*, THE VERGE (June 9, 2021), https://www.theverge.com/22522486/clearview-ai-facial-recognition-avoid-escape-privacy.

[187] *Attorney General Donovan Sues Clearview AI for Violations of Consumer Protection Act and Data Broker Law*, OFF. OF VT. ATT'Y GEN. (Mar. 10, 2020), https://ago.vermont.gov/blog/2020/03/10/attorney-general-donovan-sues-clearview-ai-violations-consumer-protection-act-and-data-broker-law.

[188] Johana Bhuiyan, *Clearview AI Uses Your Online Photos to Instantly ID You. That's A Problem, Lawsuit Says*, L.A. TIMES (Mar. 9, 2021), https://www.latimes.com/business/technology/story/2021-03-09/clearview-ai-lawsuit-privacy-violations.

opt out of the use of that information to build the Products.

166. Defendants' violation of the law is ongoing as they continue to collect personal brokered information by scraping the internet without registering as data brokers or otherwise providing notice or seeking consent from anyone. Plaintiff and the Classes have a right to opt out of this ongoing scraping of internet information but no mechanism to exercise that right, absent the injunctive relief sought in this Action.

**B. Defendants' Web Scraping Violated Plaintiff's Property Interests**

167. Courts recognize that internet users have a property interest in their personal information and data. *See Calhoun v. Google, LLC,* 526 F. Supp. 3d 605, 635 (N.D. Cal. 2021) (recognizing property interest in personal information and rejecting Google's argument that "the personal information that Google allegedly stole is not property"); *In re Experian Data Breach Litigation,* SACV 15-1592 AG (DFMx), 2016 U.S. Dist. LEXIS 184500, at *14 (C.D. Cal. Dec. 29, 2016) (loss of value of personal identifying information is a viable damages theory); *In re Marriott Int'l Inc. Customer Data Sec. Breach Litig.,* 440 F. Supp. 3d 447, 460-61 (D. Md. 2020) ("The growing trend across courts that have considered this issue is to recognize the lost property value of this [personal] information."); *Simona Opris v. Sincera,* No. 21-3072, 2022 U.S. Dist. LEXIS 94192, at *20 (E.D. Pa. May 23, 2022) (collecting cases).

168. Plaintiff' and Class Members' property rights in the personal data and information that they have generated, created, or provided through various online platforms thus includes the right to possess, use, profit, sell, and exclude others from accessing or exploiting that information without consent or renumeration.

169. The economic value of this property interest in personal information is well understood, as a robust market for such data drives the entire technology economy. As experts have noted, the world's most valuable resource is "no longer oil, but data," and has been for years now.[189]

170. A single internet user's information can be valued anywhere from $15 to $40, and

[189] *The World's Most Valuable Resource Is No Longer Oil, but Data*, THE ECONOMIST (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

even more.[190] Another study found that an individual's online identity can be sold for $1,200 on the dark web.[191] Defendants' misappropriation of every piece of data available on the internet, and with it, millions of internet users' personal information without consent, thus represents theft of a value unprecedented in the modern era of technology.

171. Writing for the Harvard Law Review, Professor Paul M. Schwartz underscored the value of personal data, as follows: "Personal information is an important currency in the new millennium. The monetary value of personal data is *large* and still *growing*, [and that's why] corporate America is moving quickly to profit from the trend."[192] The data forms a critical "corporate asset."

172. Other experts concur: "[S]uch vast amounts of collected data have obvious and substantial economic value. Individuals' traits and attributes (such as a person's age, address, gender, income, preferences… [their] clickthroughs, comments posted online, photos updated to social media, and so forth) are increasingly regarded as business assets[.]"[193]

173. Because personal data is valuable personal property, market exchanges now exist where internet users like Plaintiff and putative class members can sell or monetize their own personal data and internet usage information.[194] For example, Facebook has offered to *pay* users for their voice recordings.[195] By contrast and as alleged herein upon information and belief, Defendants

---

[190] *Id.*

[191] Maria LaMagna, *The Sad Truth About How Much Your Facebook Data is Worth on the Dark Web*, MARKETWATCH (June 6, 2018), https://www.marketwatch.com/story/spooked-by-the-facebook-privacy-violations-this-is-how-much-your-personal-data-is-worth-on-the-dark-web-2018-03-20.

[192] Paul M. Schwartz, Property, Privacy, and Personal Data, 117 HARV. L. REV. 2056, 2056 (May, 2004).

[193] Alessandro Acquisti et al., *The Economics of Privacy*, 54(2) J. OF ECON. LITERATURE 442, 444 (Mar. 8, 2016).

[194] Kevin Mercandante, *10 Apps for Selling Your Data for Cash*, BEST WALLET HACKS, https://wallethacks.com/apps-for-selling-your-data/ (last updated November 18, 2023); Kari Paul, *Facebook Launches Apps That Will Pay Users for Their Data*, THE GUARDIAN (June 11, 2019) https://www.theguardian.com/technology/2019/jun/11/facebook-user-data-app-privacy-study; Saheli Roy Choudry & Ryan Browne, *Facebook Pays Teens to Install an App That Could Collect All Kinds of Data*, CNBC (Jan. 29, 2019), https://www.cnbc.com/2019/01/29/facebook-paying-users-to-install-app-to-collect-data-techcrunch.html.

[195] Tim Bradshaw, *Facebook Offers to Pay Users for Their Voice Recordings*, FIN. TIMES (Feb. 21, 2020), https://www.ft.com/content/42f6b93c-54a4-11ea-8841-482eed0038b1.

simply *took* millions of text files, voice recordings, and facial scans from across the internet — without any consent from putative class members, much less personal remuneration to them. Theft of this nature is not only unprecedented and unjust, but also dangerous. As noted in Section II, it puts millions at risk for their likeness to be cloned to perpetrate fraud, or to embarrass or otherwise harm them.

174. Moreover, the law specifically recognizes a legal interest in unjustly earned profits based on unauthorized harvesting of personal data, and "this stake in unjustly earned profits exists regardless of whether an individual planned to sell his or her data or whether the individual's data is made less valuable."[196]

175. Defendants have been unjustly enriched by their theft of personal information as its billion-dollar AI business, including ChatGPT and beyond, was built on harvesting and monetizing Internet users' personal data. Thus, Plaintiff and the Classes have a right to disgorgement and/or restitution damages representing the value of the stolen data and/or their share of the profits Defendants earned thereon.

**C. Defendants' Web Scraping Violated Plaintiff's Privacy Interests**

176. In addition to property rights, internet users maintain privacy interests in personal information even if it is posted online, and experts agree the collection, processing, and further dissemination of this information can create distinct privacy harms.[197]

177. For example, the aggregation of collected information "can reveal new facts about a person that she did not expect would be known about her when the original, isolated data was collected."[198] Even a small subset of "public" private information can be used to harm the privacy interests of internet users. One example is when researchers analyzed public tweets to identify users with mental health issues; naturally, Twitter users did not consent or expect their data to be used in that way, to potentially reveal new, highly personal information about them.[199] If that analysis were made public, or used commercially, that would pose significant and legally cognizable privacy

[196] *In re Facebook, Inc. Internet Tracking Litigation*, 956 F.3d 589, 600 (9th Cir. 2020).
[197] Geoffrey Xiao, *Bad Bots: Regulating the Scraping of Public Information*, 34(2) HARV. J.L. & TECH., 701, 706, 732 (2021).
[198] Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 493 (2006).
[199] Xiao, *supra* note 197, at 707.

harms.

178. Another reason users retain privacy interests in their personal data on the internet, even when it is technically 'public,' is the reasonable expectation of "obscurity" *i.e.*, "the notion that when our activities or information is unlikely to be found, seen, or remembered, it is, to some degree safe."[200] Privacy experts note users' reasonable expectation that most of the Internet will simply ignore their individual posts. Moreover, "[t]he passage of time also makes information obscure: no one remembers your MySpace pictures from fifteen years ago."[201]

179. Internet users' reasonable expectations are also informed by the known transaction costs that, typically, would "prevent[] someone from collecting all your photos from every social media site you have ever used – 'just because information is hypothetically available does not mean most (or even a few) people have the knowledge and ability to access ['public' private] information.'"[202]

180. When users post information on the internet, "they do so believing that their information will be obscure and in an environment of trust" on whichever site they post. Users expect a level of privacy— they **"do not expect their information to be swept up by data scraping."** Thus, according to experts, the privacy problem with "widescale, automated collection of personal information via scraping," is that it "destroys" reasonable user expectations including the right to "obscurity" by reducing the typical transaction costs and difficulties in accessing, collecting, and understanding personal information at scale.[203]

181. Scraping therefore illegally enables the use of personal information in ways which reasonable users could not have anticipated. In respect of Defendants' surreptitious scraping at unprecedented scale, it means all items users have posted on the internet have now been collected, including their voice recordings and images – arming Defendants with the ability to create a digital clone of each internet user to anticipate and manipulate their next move. Plaintiff and the Classes did not consent to such use of their personal information. As privacy experts note, "even if a user

[200] Woodrow Hartzog, *The Public Information Fallacy*, 99 BOS. L. REV. 459, 515 (2019).
[201] Xiao, *supra* note 197, at 708-09.
[202] *Id*. at 709.
[203] *Id*.

makes the affirmative choice to make [an internet post public], she manifests an intent to participate in an obscure and trustworthy environment, not an intent to participate in data harvesting."[204]

182. Worse, Plaintiff and the Classes could not have known Defendants were collecting their personal information, because Defendants did it without notice to anyone, in violation of California law which required them to register with the state as data brokers.[205]

183. Introducing these data broker laws, the California assembly stated its intent: "[C]onsumers are generally not aware that data brokers possess their personal information, how to exercise their right to opt out, and whether they can have their information deleted, as provided by California law." Thus, "it is the intent of the Legislature to further Californians' right to privacy by giving consumers an additional tool to help control the collection and sale of their personal information by requiring data brokers to register annually with the Attorney General and provide information about how consumers may opt out of the sale of their personal information."[206]

184. "Sale" of information includes "making it available" to others for consideration, which Defendants have done by commercializing the stolen data into ChatGPT and building a billion-dollar business from it. Despite scraping information for this express purpose, Defendant OpenAI did not, and still has not, registered with the State of California as required.

185. Experts acknowledge the "serious privacy harms" inherent in the type of entirely "covert information" collection in which Defendants engaged.[207] It "undermines individual autonomy and free choice."[208] The lack of notice, including under California's data broker laws, "excludes individuals from the data collection process, making individuals feel powerless in controlling how their data is used."[209] This is not just a feeling—as described *supra*, the harm is concrete economic injury given the robust market for personal information.

186. Without notice of Defendants' scraping practices, users were also denied the ability

---

[204] *Id.* at 711.
[205] Cal. Civ. Code § 1798.99.80(d).
[206] Assemb. B. 1202, 2019-2020 Reg. Sess. (Cal. 2019) (as discussed in Xiao, *supra* note 197, at 714-715).
[207] Xiao, *supra* note 197, at 719.
[208] *Id.*
[209] *Id.*

to engage in self-help, by choosing to make obscure but technically publicly-available information private – and the lack of notice precluded users from exercising their statutory data privacy rights, such as the right to request deletion.[210] Instead, Plaintiff's and the Classes' internet histories are now embedded in Defendants' AI products with no recourse other than the damages and injunctive relief requested in this Action.

### D. Defendants' Business Practices are Offensive to Reasonable People and Ignore Increasingly Clear Warnings from Regulators

187. Defendants' mass scraping of personal data for commercialization has sparked outrage over the legal and privacy implications of Defendants' practices. Those aware of the full extent of the misappropriation are fearful and anxious about how Defendants used their "digital footprint" and about how Defendants might use all that personal information going forward. Absent the relief sought in this Action, there will be no limits on such future use. The public is also concerned about how all of their personal information might be accessed, shared, and misused *by others,* now that it is forever embedded into the large language models on which the Products run.

188. The outrage makes sense: Defendants admit the Products might evolve to act against human interests, and that regardless, they are unpredictable. Thus, by collecting previously obscure and personal data of millions and permanently entangling it with the Products, Defendants knowingly put Plaintiff and the Classes in a zone of risk that is *incalculable* — but unacceptable by any measure of responsible data protection and use.

189. The extent to which Defendants stand to profit from the unprecedented privacy risks they were willing to take—with data that is not theirs—is especially offensive to everyday people. As one explained, "Using AI as it stands right now is *normalizing the illegal mass scraping* of everyone's data regardless of their nature, just to make the top even richer and forfeit any means we have to protect our work *and who we are as humans.* This should not be encouraged and tolerated."[211] The outrage stems, in part, from this uncontestable truth: "None of this would have been possible without data – *our data* – collected and used without our permission."[212]

---

[210] *Id*. at 720.
[211] @coffeeseed, TWITTER (May 11, 2023), https://twitter.com/CoffeeSeed/status/1656634134616211461.
[212] Gal, *supra* note 5.

190. In this new era of AI, we cannot allow widescale illegal data scraping to become a commercial norm; otherwise, privacy as a fundamental right will be relegated to the dustbin of history. Underscoring the need for court intervention, AI researcher Remmelt Ellen remarked simply, "[i]llegal scraping needs to be addressed."[213]

191. The public is also troubled by the lack of just compensation for the use of their personal data. One AI large language model developer stated it plainly: "If your data is used, companies should cough up."[214] Otherwise, according to a more complete critique of the current business model, AI is just "pure primitive accumulation"—taking from the masses to enrich a few, *i.e.*, Silicon Valley tech companies and their billionaire owners.[215]

192. While the past, and ongoing, misappropriation of valuable personal information is bad enough, the Products also stand to altogether eliminate future income for millions, due to the widespread unemployment they are expected to cause over time. No one has consented to the use of their personal information to build this destabilized future of social unrest and worsening poverty for everyday people, while the pockets of OpenAI and Microsoft are lined with profit.

193. As OpenAI itself once acknowledged, albeit when still purely not-for-profit, the Company would need to fund a universal basic income (UBI) if the Products were ever developed and deployed for widespread public use, because they would eliminate so many jobs. Even now, Mr. Altman's "grand idea is that OpenAI will capture much of the world's wealth through the creation of A.G.I. and then redistribute this wealth to the people."[216] Given Defendants' sudden deployment of the Products across virtually every industry using data that was not theirs, this future should begin now, with legal or equitable redistribution of Defendants' ill-gotten gains. Others have noted that a portion of the profits generated by Defendants can be funneled back "to everyone who contributed content." This would include "basically everyone," given the scope of the initial and

---

[213] @RemmeltE, TWITTER (Apr. 10, 2023), https://twitter.com/RemmeltE/status/1645499008075407364.

[214] @yudhanjaya, TWITTER (June 9, 2023), https://twitter.com/yudhanjaya/status/1667391709679095808.

[215] Bridle, *supra* note 76.

[216] Cade Metz, *The ChatGPT King Isn't Worried, but He Knows You Might Be*, THE N.Y. TIMES (Mar. 31, 2023), https://www.nytimes.com/2023/03/31/technology/sam-altman-open-ai-chatgpt.html.

ongoing theft of personal information by Defendants.[217]

194. To avoid the unjust enrichment of Defendants, this Court sitting in equity has the power to order a "data dividend" to consumers for as long as the Products generate revenue fueled on the misappropriated data. At the very least, Plaintiff and the Classes should be personally and directly compensated for the fair market value of their contributions to the large language models on which the Products were built and thrive, in an amount to be determined by expert testimony. Fundamental principles of property law demand such compensation, and everyday people reasonably support it.[218]

195. While the property and privacy rights this Action seeks to vindicate are settled as a general matter, their application to business practices surrounding the large language models fueling AI products has not been widely tested under the law. However, just weeks ago, the FTC settled an action against Amazon, in connection with the company's illegal use of voice data to train the algorithms on which its popular Alexa product runs. That action raised many of the same type of violations alleged in this Action.

196. Announcing settlement of the action, the FTC gave a stern public warning to companies like Defendants: "Amazon is not alone in apparently seeking to amass data to refine its machine learning models; right now, with the advent of large language models, the tech industry as a whole is ***sprinting*** to do the same."[219] The settlement, it continued, was to be a message to all: "Machine learning is ***no excuse to break the law***... The data you use to improve your algorithms must be ***lawfully collected*** and ***lawfully retained***. Companies would do well to heed this lesson."[220]

197. The FTC's warning comports with FTC Commissioner Rebecca Slaughter's earlier warning, in 2021, in the Yale Journal of Law and Technology.[221] Discussing the FTC's new practice

---

[217] *Id.*

[218] *See*, *e.g.*, @ianfinlay2000, *Time to Get Paid For Our Data?*, REDDIT (2021), https://www.reddit.com/r/Futurology/comments/qknz3u/time_to_get_paid_for_our_data/ ("[T]he companies are basically stealing our data bc no one knows that they should be getting paid for it").

[219] Devin Coldewey, *Amazon Settles with FTC for $25M After 'Flouting' Kids' Privacy and Deletion Requests*, TECHCRUNCH (May 31, 2023), https://techcrunch.com/2023/05/31/amazon-settles-with-ftc-for-25m-after-flouting-kids-privacy-and-deletion-requests/ (emphasis added).

[220] *Id.* (emphasis added).

[221] Rebecca Kelly Slaughter et al., *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission*, 23 YALE J. L. & TECH. 1, 39 (Aug. 2021).

of ordering "algorithmic destruction," Commissioner Slaughter explained that "the premise is simple: when companies collect data illegally, they should not be able to profit from either the data or any algorithm developed using it."[222] Commissioner Slaughter believed this enforcement approach would "send a clear message to companies engaging in illicit data collection in order to train AI models: ***Not worth it.***"[223] Unfortunately for the millions of consumers impacted by Defendants' mass theft of data, Defendants did not heed the warning.

**E. Defendants' Theft of User Data in Excess of Reasonable Consent**

198. Defendants' second category of theft stems from their unrestricted harvesting of data from Users of the Products, including registered Users of the OpenAI website and Users of Defendants' API and/or plug-ins.

199. Defendants have made much of the fact that they purportedly "want" to comply with applicable privacy laws and regulations—and will likely oppose this lawsuit by arguing that registered users of the Products purportedly "consented" to the widespread theft of their personal information by virtue of using the Products. This argument is disingenuous for multiple reasons.

200. *First*: For those consumers who used ChatGPT plug-ins or API, the various sites' use policies did not provide anything approaching informed consent that the consumers' information and personal data would be used to train Defendants' LLMs and would thus be incorporated into generative AI in a manner that would prevent them from reasonably ever removing their data from Defendants' for-profit commercial enterprises. Plaintiff and Class Members had no idea that Defendants were and are collecting and utilizing their User Data, including the most sensitive information, when they engage with ChatGPT which seamlessly incorporated artificial intelligence in the background.

201. Plaintiff fell victim to Defendants' unlawful collection and sharing of their sensitive information acquired through their interactions with Defendants' Products and websites, as well as the hundreds or thousands of applications that now use ChatGPT-based plug-ins or API.[224]

---

[222] *Id.*

[223] *Id.* (emphasis added).

[224] Matt Burgess, *ChatGPT Has a Big Privacy Problem*, WIRED (Apr. 4, 2023), https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/.

202. In less than 24 hours after Defendants announced the ability to install plug-ins to ChatGPT, many companies immediately jumped on board and started incorporating their websites within the AI plug-in. In exchange, Defendants received yet another wealth of personal data, once again, without the users' and nonusers' consent. ChatGPT is becoming the single app "to rule them all."[225]

203. Defendants' AI has become the virtual spy,[226] closely monitoring, recording, and training on the personal data, clicks, searches, inputs, and personal information of millions of unsuspecting individuals who may be using an Instacart to purchase grocery items, a telehealth company to make a doctor's appointment, or simply browsing Expedia to make vacation plans.

204. *Second*: Even those who registered for OpenAI accounts and interacted with ChatGPT directly did not give effective consent for Defendants to use their data and personal information in the way they currently do.

205. For instance, when Plaintiff logged in to use the ChatGPT, Defendants were tracking and collecting every piece of information entered into the chatbot—including sensitive information such as (1) all details entered into the chatbot; (2) account information users enter when signing up; (3) name; (4) contact details; (5) login credentials; (6) emails; (7) payment information; (8) transaction records; (9) identifying data ChatGPT pulls from users' devices or browsers, like IP addresses and locations; (10) social media information; (11) chat log data; (12) usage data; (13) analytics; and (14) cookies. However, Defendants are also tracking the information from other applications in which their AI is already plugged in – Stripe, Microsoft Teams, Bing, Zillow, Expedia, Instacart, etc. – and using each piece of information to train the AI.

206. Plaintiff, and all Class Members, did not consent to such extensive collection of data, and the use of their data for essentially any purpose to benefit Defendants' businesses – including for training purposes of the AI. In fact, Plaintiff and all Class Members could not consent to Defendants' conduct because they were unaware their sensitive information would be collected and

[225] Better Product, *OpenAI's Master Plan to Turn ChatGPT into an Everything App*, MEDIUM (Mar. 25, 2023), https://medium.com/@betterproducts/openais-master-plan-to-turn-chatgpt-into-an-everything-app-1270686074f8.

[226] *Id.*

used in this manner in the first place. Thus, Defendants did not obtain *valid enforceable* consent to collect, use, and store Plaintiff's and Class Members' sensitive information.

207. In the near future, Defendants anticipate adding even more powerful features to the omniscient AI, allowing it to also gather data from audio inputs with their yet another AI—Vall-E. Vall-E has already been developed and allows to process three (3) seconds of a human voice, and be able to speak in such voice in perpetuity. Once activated, Defendants' and their AI's access to human voices and audio inputs will jeopardize the users' and nonusers' privacy even further.

208. Defendant OpenAI has also deceptively represented to its users that they can request their private information not be used and, if parents discover that a child has used ChatGPT, Defendant will erase the child's data from the system. This is deceptive because by the time the language model has taken in the information and learned from it, that information has already financially benefited Defendants and cannot be removed from the knowledge base of the language model. Moreover, Defendant OpenAI has stated that, notwithstanding a user's requests to opt out of data collection and sharing, it will still retain some information (though what information will be retained is not specified).

209. Currently, a ChatGPT user wanting to opt out of the use of their data and chats for model training is instructed that they can simply turn off chat history (which deprives them of using that functionality themselves) and the application will stop using *new* chat content for training purposes.[227] However, Defendants continue to train their models with the user's information – be it from the prior chats or new chats. Moreover, as noted above, it is impossible to know whether any of the previously used data can effectively be deleted, as once the language model is trained using the data, it becomes part of the model. Additionally, the option of opting out of chat history retention doesn't impact OpenAI's ability to use a user's other personal data gathered during the account creation process for Defendants' own purposes. OpenAI's privacy disclosures are intentionally vague about this, noting simply that a user can opt out of chat history retention *or* can submit a form to ask OpenAI not to use or share their data. No guidelines are provided regarding whether or when

[227] Johanna C., *How Do I Turn Off Chat History and Model Training?*, OPENAI, https://help.openai.com/en/articles/7792795-how-do-i-turn-off-chat-history-and-model-training (last visited February 14, 2024).

Defendant might decline to honor such a request, nor how long it takes to process.

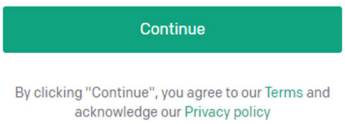
210. Furthermore, as commentators have observed, Defendant OpenAI heavily pushes users not to opt out of data collection.[228] Once a user turns off the option for their ChatGPT interactions to be used for training purposes, they are presented constantly with a large green button that encourages them to "Enable chat history." Nothing on this button notifies users that enabling chat history functionality amounts to reauthorizing OpenAI to save and train Defendants' models on the user's data.

211. Moreover, it is not clear what information a given user can actually prevent OpenAI from retaining and using in the future, as the company has stated in blog posts that it will retain some data anyway and that some of this data can be used in Defendant OpenAI's training datasets.[229]

212. Defendants fail to provide accurate and comprehensive notifications to consumers about the scale of their data sharing practices. Defendants' admissions within their Privacy Policy do not adequately inform consumers on the breadth of data sharing, resulting in a breach of explicit assurances and a violation of reasonable consumer expectations. By acting in such a manner, Defendants are engaged in data misuse practices that contradict the principles of transparency, accountability, and respect for consumer privacy rights.

***1. OpenAI's disclosures are not conspicuous.***

213. When a consumer attempts to register for an OpenAI account, they are presented with the following image:




214. When a hyperlink to an agreement is "not conspicuous enough to put [Plaintiff] on inquiry notice," then the agreement is not binding. *Colgate v. JUUL Labs, Inc.*, 402 F. Supp. 3d

[228] Natasha Lomas, *How to Ask OpenAI for Your Personal Data to Be Deleted or Not Used to Train Its AIs*, TECHCRUNCH (May 2, 2023), https://techcrunch.com/2023/05/02/chatgpt-delete-data/.

[229] Yaniv Markovski, *How Your Data Is Used to Improve Model Performance*, OPENAI, https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance (last visited February 14, 2024).

728, 764-66 (N.D. Cal. 2019). The Ninth Circuit holds that "even close proximity of the hyperlink to relevant buttons users must click on—without more—is insufficient to give rise to constructive notice." *Nguyen v. Barnes & Noble Inc.*, 763 F.3d 1171, 1179 (9th Cir. 2014). Instead, courts consider factors such as color, size and font of the hyperlink, and whether the hyperlink is presented alone or in a clutter of text. *See, e.g.*, *Colgate*, 402 F. Supp. 3d at 764; *Selden v. Airbnb, Inc.*, 16-cv-00933 (CRC), 2016 WL 6476934, at *14-15 (D.D.C. Nov. 1, 2016).

215. Here, a consumer registering for an OpenAI account is ferried through the process and is provided only small hyperlinks to OpenAI's Privacy Policy and Terms of Use during the sign-up process. The lettering alerting the potential registrant to the documents is tiny and gray. The consumer need not make any indication that he or she has actually read the documents, nor that they understand the connection between these documents and their creation of an account. Unlike many companies that require a consumer to scroll to the bottom of a privacy policy or other legal document—or at least click a radial purporting to have read the document—an OpenAI registrant need make no affirmative indication that they are aware of the policies whatsoever. As such, there is no binding agreement between Defendant OpenAI and Plaintiff or the Members of the Subclasses regarding use of these individuals' information, and no effective consent.

216. Plaintiff and the User Subclasses were neither on constructive notice nor inquiry notice of the privacy policy on the ChatGPT platform.

### *2. Defendants' Use of Consumer Data Far Exceeds Industry Standards and their Own Representations*

217. The Federal Trade Commission has promulgated numerous guides for businesses highlighting the importance of implementing reasonable data security practices. According to the FTC, the need for data security should be factored into all decision-making.[230]

218. In 2016, the FTC updated its publication, Protecting Personal Information: A Guide for Business, which established cybersecurity guidelines for businesses.[231] The guidelines note that

[230] *Start with Security: A Guide for Business: Lessons Learned from FTC Cases*, FED. TRADE COMM'N. (June 2015), https://www.ftc.gov/system/files/documents/plain-language/pdf0205-startwithsecurity.pdf.

[231] *Protecting Personal Information: A Guide for Business*, FED. TRADE COMM'N. (Oct. 2016), https://www.ftc.gov/system/files/documents/plain-language/pdf-0136_proteting-personal-information.pdf.

businesses should protect the personal customer information that they keep; properly dispose of personal information that is no longer needed; encrypt information stored on computer networks; understand their network's vulnerabilities; and implement policies to correct any security problems.

219. The FTC further recommends that entities not maintain personally identifiable information longer than is needed for authorization of a transaction; limit access to sensitive data; require complex passwords to be used on networks; use industry-tested methods for security; monitor for suspicious activity on the network; and verify that third-party service providers have implemented reasonable security measures. The FTC has brought enforcement actions against entities engaged in commerce for failing to adequately and reasonably protect customer data, treating the failure to employ reasonable and appropriate measures to protect against unauthorized access to confidential consumer data as an unfair act or practice prohibited by Section 5 of the Federal Trade Commission Act ("FTCA"), 15 U.S.C. § 45. Orders resulting from these actions further clarify the measures businesses must take to meet their data security obligations.

220. Defendants fail to meet these obligations, as they directly feed consumers' personal information into their LLMs for training purposes.

221. Even if the click-through button discussed above could constitute a binding agreement—it cannot—the substance of the policies is insufficient to put any consumer on notice of what to expect with regard to the use of their information. The policies lay out vague promises regarding how and when the users' data can and will be shared, and affirm that all laws are being complied with—even where such affirmations are internally inconsistent.[232] For example, under the heading "Additional U.S. State Disclosures," the Privacy Policy lists five different categories of "Personal Information," including one category that OpenAI identifies as "Sensitive Personal Information," and states that OpenAI discloses information from *all five* of the various categories to "our affiliates, vendors and service providers, law enforcement, and parties involved in Transactions." Yet a few paragraphs down, the policy then inexplicably asserts "We don't sell Personal Information or share Personal Information." No explanation is given as to what is meant

[232] *Privacy Policy*, OPENAI, https://openai.com/policies/privacy-policy (last updated November 14, 2023).

by the assertion that the company both *does* and *does not* share Personal Information.

222. As of June 23, 2023, Defendants changed this language to clarify that they "don't 'sell' Personal Information or 'share' Personal Information for cross-contextual behavioral advertising (as those terms are defined under applicable local law)."[233] Nevertheless, no explanation is given as to how Defendants can ensure that the entities with which they are sharing users' personal information with are not, in fact, using it for cross-contextual behavior advertising. Defendants also do not disclose the specific purposes for which they do use such sensitive data.

223. Moreover, the Policy alerts consumers that to the extent local law entitles them to request deletion of their Personal Information, they can exercise this right (amongst others) by sending a request to dsar@openai.com. Yet nothing in the privacy policy explains that information which has already been incorporated into Defendants' LLMs *can never really* be removed.

224. Finally, even if users are on notice of the Privacy Policy (and they are not), the Privacy Policy does not disclose wiretapping. There is **zero** adequate consent for wiretapping, and OpenAI's terms and conditions are convoluted, inconspicuous, and consist of numerous documents, impossible to decipher by reasonable consumers. There are no conspicuous or clear disclosures that all conversations are wiretapped, recorded, and shared with numerous entities—none of which are disclosed.

225. Beyond Defendants' legal obligations to protect the confidentiality of individuals' User Data, Defendants' privacy policy and online representations affirmatively and unequivocally state that any personal information provided to Defendants will remain secure and protected. Since ChatGPT's inception, Defendants have represented and continue to represent that:

> "We at OpenAI OpCo, LLC (together with our affiliates, "OpenAI", "we", "our" or "us") respect your privacy and are strongly committed to keeping secure any information we obtain from you or about you."
>
> "We implement commercially reasonable technical, administrative, and organizational measures to protect Personal Information both online and offline from loss, misuse, and unauthorized access, disclosure, alteration, or destruction."
>
> "OpenAI does not knowingly collect Personal Information from children

[233] *Id.*

under the age of 13."[234]

226. Defendants have failed to adhere to a single promise vis-à-vis their duty to safeguard User Data. Defendants have made these privacy policies and commitments available in ChatGPT. In these representations to Plaintiff and Class Members and the public, Defendants promised to take specific measures to protect its members' information, consistent with industry standards and federal and state law. However, they did not.

227. Plaintiff and Class Members relied to their detriment on Defendants' uniform representations and omissions regarding data security. Now that their sensitive personal and medical information is in the possession of third parties, Plaintiff and Class Members face a constant threat of continued harm. Collection of such sensitive information without consent or notice poses a great threat to individuals by subjecting them to the danger of potential attacks and embarrassment.

228. Plaintiff and Class Members trusted Defendants' Products when inputting sensitive and valuable User Data. Had Defendants disclosed to Plaintiff and its other members that every click, every search, and every input of sensitive information was being tracked, recorded, collected, and disclosed to third parties—Plaintiff would not have trusted Defendants' Products to input such sensitive information.

229. Defendants knew or should have known that Plaintiff and Class Members would reasonably rely upon, and trust Defendants' promises regarding security and safety of its data and systems.

230. Additionally, Defendants were aware that ChatGPT collects, tracks, and discloses Plaintiff's and Class Members' User Data, including sensitive information.

231. By virtue of how ChatGPT is "trained," *i.e.*, through the collection and processing of a massive corpus of data, Defendants were aware that their Users' data would be collected and disclosed to third parties every time a user interacted with ChatGPT.

## CLASS ALLEGATIONS

232. **Class Definition:** Plaintiff brings this action pursuant to Federal Rules of Civil Procedure Sections 23(b)(2), 23(b)(3), and 23(c)(4), on behalf of Plaintiff and the Class defined as

[234] *Id.*

follows:

a. **Non-User Class**: All persons in the United States whose PII, Personal Information, or Private Information was disclosed to, or accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

b. **ChatGPT User Class:** All persons in the United States who used ChatGPT, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

c. **ChatGPT API User Class:** All persons in the United States who used other platforms, programs, or applications which integrated ChatGPT technology, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

d. **Microsoft User Class:** All persons in the United States who used Microsoft platforms, programs, or applications which integrated ChatGPT technology, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

e. **ChatGPT Plus User Class:** All persons in the United States who used Chat-GPT website or mobile app and whose Personal Information or PII was intercepted, accessed, collected, tracked, stored, shared, taken, or used by Defendants without consent and/or authorization.

**State-Wide Subclasses:**

**The California Subclasses**

i. **California Non-User SubClass**: All persons within the State of California whose PII, Personal Information, or Private Information was disclosed to, or accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

ii. **California ChatGPT User SubClass:** All persons within the State of California who used ChatGPT, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

iii. **California ChatGPT Plus User SubClass:** All persons within the State of California who used ChatGPT website or mobile app and whose Personal Information or PII was intercepted, accessed, collected, tracked, stored, shared, taken, or used by Defendants without consent and/or authorization.

**The New York Subclasses**

i. **New York Non-User SubClass**: All persons within the State of New York whose PII, Personal Information, or Private Information was disclosed to, or accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

ii. **New York ChatGPT User SubClass:** All persons within the State of New York who used ChatGPT, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or

used by Defendants without consent or authorization.

iii. **New York ChatGPT Plus User SubClass:** All persons within the State of New York who used ChatGPT website or mobile app and whose Personal Information or PII was intercepted, accessed, collected, tracked, stored, shared, taken, or used by Defendants without consent and/or authorization.

233. **The following people are excluded from the Classes and Subclasses:** (1) any Judge or Magistrate presiding over this action and members of their judicial staff and immediate families; (2) Defendants, Defendants' subsidiaries, parents, successors, predecessors, and any entity in which the Defendants or their parents have a controlling interest and its current or former officers and directors; (3) persons who properly execute and file a timely request for exclusion from the Class; (4) persons whose claims in this matter have been finally adjudicated on the merits or otherwise released; (5) Plaintiff's counsel and Defendants' counsel; and (6) the legal representatives, successors, and assigns of any such excluded persons.

234. Plaintiff reserves the right under Federal Rule of Civil Procedure 23 to amend or modify the Class to include a broader scope, greater specificity, further division into subclasses, or limitations to particular issues. Plaintiff reserves the right under Federal Rule of Civil Procedure 23(c)(4) to seek certification of particular issues.

235. The requirements of Federal Rules of Civil Procedure 23(a), 23(b)(2), and 23(b)(3) are met in this case.

236. The Fed. R. Civ. P. 23(a) elements of Numerosity, Commonality, Typicality, and Adequacy are all satisfied.

237. **Ascertainability:** Membership of the Classes and Subclasses is defined based on objective criteria and individual members will be identifiable from Defendants' records, records of third-party platforms/applications which integrate ChatGPT, including the massive data storage, consumer accounts, and enterprise services that Defendants offer. Identification is also available through self-identification methods.

238. **Numerosity:** The precise number of the Members of Classes and Subclasses is not available to Plaintiff, but individual joinder is demonstrably impracticable.

239. **Commonality:** Commonality requires that the Members of Classes and Subclasses

allege claims which share common contention such that determination of its truth or falsity will resolve an issue that is central to the validity of each claim in one stroke. Here, there is a common contention for all Classes and Subclasses as follows:

**Defendants' Web-Scraping Practices (Non-User Class)**

a) Whether the members of Non-User Class had a protected property right in their data;

b) Whether Defendants scraped the protected data belonging to Non-User Class Members without consent;

c) Whether Defendants' collection, scraping, and uses of the protected Non-User Class Members of protected data violates:

1. Electronic Communication Privacy Act, 18 U.S.C. §§ 2510, et. seq.
2. Computer Fraud and Abuse Act, 18 U.S.C. §§ 1030, et. seq.
3. California Constitution right to privacy;
4. California Invasion of Privacy Act, Cal. Pen. Code §§ 630, et seq.
5. California Unfair Competition Law, Bus. & Prof Code § 17200;
6. New York General Business Law §§ 349, et seq.

d) Whether Defendants' collection, scraping, and uses of the protected Non-User Class Members of protected data constitutes:

1. Common law Negligence;
2. Unlawful Intrusion upon Seclusion under California laws;
3. Conversion;
4. Larceny/Receipt of Stolen Property under Cal. Pen. Code § 496(a) and (c).

e) Whether as a result of Defendants' collection, scraping, and uses of the protected Non-User Class Members of protected data, Non-User Class Members suffered monetary damages, including but not limited to actual damages, statutory damages, punitive damages, treble damages, or other monetary damages.

f) Whether as a result of Defendants' collection, scraping, and uses of the protected Non-User Class Members of protected data, Non-User Class Members are entitled

to equitable relief, including but not limited to restitution, disgorgement of profits, injunctive and declaratory relief, or other equitable remedies.

**Defendants' Collection/Interception Practices of Private Information From ChatGPT User, ChatGPT Plug-In User, ChatGPT Plus User Classes, and Subclasses:**

a) Whether Defendants failed to advise the members of Classes and Subclasses the extent to which Defendants intercepted, received, or collected Private Information;

b) Whether Defendants intercepted, received, or collected communications, tracked all activities, chat history, and other Private Information from the Users of Other Platforms Which Integrate ChatGPT without consent of such Users.

c) Whether Microsoft Defendant intercepted, received, or collected communications, tracked all activities, chat history, and other Private Information of ChatGPT Users, without consent of such Users;

d) Whether OpenAI Defendant aided, abetted, and otherwise conspired with Microsoft Defendant, to allow Defendant Microsoft's interception, receipt, or collection of communications, tracking of all activities, and other Private Information of ChatGPT Users, without consent of such Users;

e) Whether Defendants' conduct of intercepting, receipt, or collection of Private Information of the members of Classes and Subclasses violated federal and state privacy laws, anti-wiretapping laws, or other tort laws, including but not limited to:

1. Electronic Communication Privacy Act, 18 U.S.C. § 2510 *et. seq.*
2. Computer Fraud and Abuse Act, 18 U.S.C. § 1030 *et. seq.*
3. California Constitution right to privacy;
4. California Invasion of Privacy Act, Cal. Pen. Code §§ 630 et seq.
5. California Unfair Competition Law, Bus. & Prof Code §§ 17200;
6. Common law Negligence;
7. Unlawful Intrusion upon Seclusion under California laws;
8. Conversion.

f) Whether as a result of Defendants' collection, scraping, and uses of the protected

Private Information, ChatGPT User, ChatGPT Plug-In User, or ChatGPT Plus User Class Members and Subclass Members suffered monetary damages, including but not limited to actual damages, statutory damages, punitive damages, treble damages, or other monetary damages.

g) Whether as a result of Defendants' interception, collection, receipt, or unauthorized uses of Private Information, ChatGPT User, ChatGPT Plug-In User, or ChatGPT Plus User Class Members and Subclass Members are entitled to equitable relief, including but not limited to restitution, disgorgement of profits, injunctive and declaratory relief, or other equitable remedies.

240. **Typicality:** Plaintiff's claims are typical of the claims of other Class Members in that Plaintiff and the Class Members sustained damages arising out of Defendants' uniform wrongful conduct and data collecting practices, interception/sharing of the collected data with each other, and use of such data in attempt to train the AI Products, and further develop the Products.

241. **Adequate Representation:** Plaintiff will fairly and adequately represent and protect the interests of the Members of Classes and Subclasses. Plaintiff's claims are made in a representative capacity on behalf of the Members of Classes and Subclasses. Plaintiff has no interests antagonistic to the interests of the other Members of Classes and Subclasses. Plaintiff has retained competent counsel to prosecute the case on behalf of Plaintiff and the Class. Plaintiff and Plaintiff's counsel are committed to vigorously prosecuting this action on behalf of the Members of Classes and Subclasses.

242. **This case also satisfies Fed. R. Civ. P. 23(b)(3) - Predominance:** There are many questions of law and fact common to the claims of Plaintiff and Members of Classes and Subclasses, and those questions predominate over any questions that may affect individual Class Members. Common questions and/or issues for Class members include the questions listed above in *Commonality,* and also include, but are not necessarily limited to the following:

a) Whether Defendants violated the California Invasion of Privacy Act;

b) Whether Defendants' unauthorized disclosure of Users' sensitive information was negligent;

c) Whether Defendants owed a duty to Plaintiff and Class Members not to disclose their sensitive user information to unauthorized third parties;

d) Whether Defendants breached their duty to Plaintiff and Class Members not to disclose their sensitive user information to unauthorized third parties;

e) Whether Defendants represented to Plaintiff and Class Members that they would protect Plaintiff and the Members of Classes and Subclasses Private Information;

f) Whether Defendants violated Plaintiff's and Class Members' right to privacy;

g) Whether Plaintiff and Class members are entitled to actual damages, enhanced damages, statutory damages, restitution, disgorgement, and other monetary remedies provided by equity and law;

h) Whether Defendants' conduct was unlawful or deceptive;

i) Whether Defendants were unjustly enriched by their conduct under the laws of California.

j) Whether Defendants fraudulently concealed their conduct; and

k) Whether injunctive and declaratory relief and other equitable relief is warranted.

243. **Superiority:** This case is also appropriate for class certification because class proceedings are superior to all other available methods for the fair and efficient adjudication of this controversy as joinder of all parties is impracticable. The damages suffered by individual Members of Classes and Subclasses will likely be relatively small, especially given the burden and expense of individual prosecution of the complex litigation necessitated by Defendants' actions. Thus, it would be virtually impossible for the individual Members of Classes and Subclasses to obtain effective relief from Defendants' misconduct. Even if Class Members could mount such individual litigation, it would still not be preferable to a class action, because individual litigation would increase the delay and expense to all parties due to the complex legal and factual controversies presented in this Complaint. By contrast, a class action presents far fewer management difficulties and provides the benefits of single adjudication, economy of scale, and comprehensive supervision by a single Court. Economies of time, effort, and expense will be enhanced, and uniformity of decisions ensured.

244. Likewise, particular issues under Rule 23(c)(4) are appropriate for certification because such claims present only particular, common issues, the resolution of which would advance the disposition of this matter and the parties' interests therein.

**CALIFORNIA LAW SHOULD APPLY TO OUT-OF-STATE PLAINTIFF'S & CLASS MEMBERS' NON-STATUTORY CLAIMS**

245. Courts "have permitted the application of California law where the plaintiff's claims were based on alleged misrepresentations [or misconduct] that were disseminated from California." *Ehret v. Uber Technologies, Inc.,* 68 F. Supp. 3d 1121, 1130 (N.D. Cal. 2014). "California courts have concluded that state statutory remedies may be invoked by out-of-state parties when they are harmed by wrongful conduct occurring in California." *In re iPhone 4S Consumer Litig.*, No. C 12-1127 CW, 2013 WL 3829653, at *7 (N.D. Cal. July 23, 2013) (internal quotation marks and citation omitted).

246. This is particularly true for non-statutory claims where the defendant has a choice-of-law provision that applies California law to that defendant's conduct.

247. However, there is sound public policy to allow statutory claims from other states to proceed against a defendant regardless of that defendant's choice of law provision. *See, e.g.*, *In re Facebook Biometric Info. Priv. Litig.*, 185 F. Supp. 3d 1155, 1168–70 (N.D. Cal. 2016).

248. Defendant OpenAI is headquartered in California; this is where Defendant OpenAI's nerve center of its business operations is located. This is where Defendant OpenAI has its high-level officers direct, control, coordinate, and manage its activities, including policies, practices, research and development, and other decisions affecting Defendants' Products. This is where the majority of unlawful conduct took place – from development of the AI products, decisions concerning AI Products and training of the AI, web scraping practices, and other major decisions which affected all Class Members. Furthermore, Defendant Microsoft operates in the state of California. Upon information and belief, decisions concerning Defendants' Products were entered into in California.

249. Furthermore, Defendant OpenAI requires that California law applies to disputes between Defendant OpenAI and ChatGPT Users.

250. The State of California, therefore, has significant interests to protect all residents and citizens of the United States against a company headquartered and doing business in California, and

has a greater interest in the claims of Plaintiff and the Classes than any other state, and the state most intimately concerned with the claims and outcome of this litigation.

251. California has significant interest in regulating the conduct of businesses operating within its borders, and that California has the most significant relationship with Defendants – as Defendant OpenAI is headquartered in California, and Defendant Microsoft conducts business (at least as it relates to Defendant OpenAI) in California, there is no conflict in applying California law to non-resident consumer claims.

252. Excluding out-of-state statutory claims, application of California law to the Classes' claims is neither arbitrary nor fundamentally unfair because choice of law principles applicable to this action support the application of California law to the nationwide claims of all Class Members.

253. Application of California law to Defendants is consistent with constitutional due process.

**COUNT ONE: VIOLATION OF ELECTRONIC COMMUNICATIONS PRIVACY ACT, 18 U.S.C. § 2510, *et seq.***
**(on behalf of ChatGPT, ChatGPT API User, and Microsoft User Classes against Defendants)**

254. Plaintiff hereby incorporates Paragraphs 1 through 253 as if fully stated herein.

255. The Federal Wiretap Act, as amended by the Electronic Communications Privacy Act of 1986 (the "Wiretap Act"), prohibits the intentional interception of the contents of any wire, oral, or electronic communication through the use of a device. 18 U.S.C. § 2511.

256. The following constitute "devices" within the meaning of the Wiretap Act, 18 U.S.C. § 2510(5):

a. The computer codes and programs that Defendants use to track the Plaintiff's and Class members' communications;

b. The Plaintiff's and Class members' browsers and applications;

c. The Plaintiff's and Class members' computing and mobile devices;

d. Defendants' web servers;

e. The web servers of websites from which Defendants tracked and intercepted the Plaintiff's and Class members' communications;

f. The computer codes and programs used by Defendants to effectuate their

tracking and interception of the Plaintiff's and Class members' communications;

g. The plan that Defendants carried out to effectuate its tracking and interception of the Plaintiff's and Class members' communications.

257. The Wiretap Act protects both the sending and reception of communications.

258. The Wiretap Act provides a private right of action to any person whose wire, oral, or electronic communication is intercepted. 18 U.S.C. § 2520(a).

259. Defendants' actions in tracking and intercepting users' communications were intentional. On information and belief, Defendants are aware that they are tracking and intercepting these communications as outlined in this complaint and they have taken no remedial actions.

260. Defendants' actions were done contemporaneously with the Plaintiff's and Class members' sending and receiving those communications.

261. Defendants' interception included "contents" of electronic communications made from Plaintiff and Class members to websites and other web properties other than Defendants' in the form of detailed URL requests, webpage browsing histories, search queries, and other information that Plaintiff and Class members sent to those websites and for which Plaintiff received communications in return from those websites.

262. The transmission of data between Plaintiff and Class members on the one hand and the websites and other web properties other than Defendants' on which Defendants tracked and intercepted Plaintiff's and Class members' communications on the other, without authorization were "transfer[s] of signs, signals, writing, . . . data, [and] intelligence of [some] nature transmitted in whole or in part by a wire, radio, electromagnetic, photoelectronic, or photooptical system that affects interstate commerce[,]" and therefore qualify as "electronic communications" within the meaning of the Wiretap Act. 18 U.S.C. § 2510(12).

263. Defendants, in their conduct alleged herein, were not providing an "electronic communication service," as that term is defined in 18 U.S.C. § 2510(12) and is used elsewhere in the Wiretap Act. Defendants were not acting as an Internet Service Provider and the conduct alleged herein does not arise from their provision of separate lines of business.

264. None of the Defendants were authorized parties to the communications because Plaintiff and Class members were unaware of the collection and interception. Neither can Defendants manufacture their own status as parties to the communications by surreptitiously intercepting those communications.

265. Defendants had a tortious and/or criminal intent in (a) obtaining the Private Information, (b) sharing the Private Information with each other; (c) feeding the Private Information into the Products, to train, develop, and commercialize their Products. Their actions were knowing and deliberate, especially since Defendants were well aware that consumers did not want nor allow Defendants to use their Private Information for training of the Products.

266. **Electronic Communications**. Electronic communication means any "transfer[s] of signs, signals, writing, . . . data, [and] intelligence of [some] nature transmitted in whole or in part by a wire, radio, electromagnetic, photoelectronic, or photooptical system that affects interstate commerce." 18 U.S.C. § 2510(12). Here, the following communications qualify as "communications" under the ECPA:

a) **Communications On ChatGPT**: Plaintiff's and Class Members' communications (including but not limited to chats, comments, replies, searches, keystrokes, signals, mouse clicks, or other data, activity, or intelligence) on ChatGPT intercepted by Defendant Microsoft;

b) **ChatGPT Intercepted Communications On Platforms Which Integrated ChatGPT API**: Plaintiff's and Class Members' communications (including but not limited to chats, comments, replies, searches, keystrokes, signals, mouse clicks, or other data, activity, or intelligence) on various applications, platforms, or websites which integrate ChatGPT API (*i.e.* Stripe, Snapchat, etc.) intercepted by Defendants;

c) **Communications on Microsoft Platforms**: Plaintiff's and Class Members' communications (including but not limited to chats, comments, replies, searches, keystrokes, mouse clicks, signals, or other data, activity, or intelligence) on Microsoft platforms which integrate ChatGPT API (*i.e.* Microsoft Teams, Outlook, etc.) intercepted by Defendant OpenAI;

267. **Content**. The ECPA defines content, when used with respect to electronic communications, to "include [] any information concerning the substance, purport, or meaning of that communication." 18 U.S.C. § 2510(8).

268. Plaintiff, and the members of all Classes and Subclasses have an expectation of privacy in their communications, entered keystrokes, chats, comments, replies, searches, signals, and other data, activity, or intelligence, and they exercised a reasonable expectation of privacy concerning the transmission of that content.

269. **Interception**. The ECPA defines interception as the "acquisition of the contents of any wire, electronic, or oral communication through the use of any electronic, mechanical, or other device" and "contents . . . include [] any information concerning the substance, purport, or meaning of that communication." 18 U.S.C. §§ 2510(4), (8).

270. Defendants intentionally accessed, and obtained access to the contents of Plaintiff's, the Classes', and Subclasses' protected computers and obtained information concerning the substance, purport, or meaning of communications, thereby, and in doing so, exceeded authority granted by Plaintiff, the Classes, and Subclasses to access the protected computers.

271. **Electronic Communication Service.** The ECPA defines electronic communication service as "any service which provides to users thereof the ability to send or receive wire or electronic communications." 18 U.S.C § 2510(15). The following services constitute "electronic communication services:"

(1) Reddit, Twitter, YouTube, Spotify, TikTok, and other websites which were scraped by Defendants;

(2) Third Party websites, programs, and applications, which integrate ChatGPT technology;

(3) Microsoft platforms, programs, applications, and websites, which integrate ChatGPT technology;

(4) OpenAI website and mobile application(s) for ChatGPT.

272. **Electronic, Mechanical, or Other Device**. The ECPA defines "electronic, mechanical, or other device" as "any device...which can be used to intercept a[n]...electronic communication[.]" 18 U.S.C. § 2510(5). The following constitute "devices" within the meaning of 18 U.S.C. § 2510(5):

(1) Plaintiff's, lasses', and Subclasses' computing devices (Mac and Windows devices

present on computers, mobile phones, tablets, or other devices);

(2) Plaintiff's, Classes', and Subclasses' browsers;

(3) Defendants' web-servers, platforms, and applications;

(4)Third-Party web-servers, platforms, and applications, where ChatGPT API technology was implemented;

(5) The tracking codes deployed by Defendants to effectuate the sending and acquisition of communications.

**I. Interception of Communications Between ChatGPT API Class Members which occurred on Third-Party Websites, Platforms, Applications, or Programs which have integrated ChatGPT API. [Microsoft User Class is Excluded]**

273. The allegations for violation of 18 U.S.C. § 2510 arising out of Defendants' interception of Plaintiff's and ChatGPT API Class Members' (collectively referred to as ChatGPT API Class Members) communications which occurred on various applications, platforms, and websites which integrate ChatGPT technology (*i.e.*, Stripe, Snapchat, etc.).

274. The transmissions of Plaintiff's and ChatGPT API Class Members' communications (including but not limited to chats, comments, replies, searches, keystrokes, mouse clicks/movements, signals, browser activity, or other data, activity, or intelligence) on various applications, programs, platforms, and websites which integrate ChatGPT technology (*i.e.*, Stripe, Snapchat, etc.) qualify as "communications" under 18 U.S.C. § 2510(12).

275. By integrating ChatGPT technology on third party platforms, Defendants are in the unique position of having unrestricted, real-time access to the users' every input, move, mouse click, chat, comment, reply, search, keystroke, browser activity, or other data, activity, or intelligence on the third-party platform.

276. As Plaintiff and ChatGPT API Class Members interact with each other or the third-party entities, Defendants intentionally tap, electrically or otherwise intercept, the lines of internet communications between Plaintiff and ChatGPT API Class Members, and/or third-party entities.

277. In disregard for Plaintiff's and ChatGPT API Class Members' privacy rights, Defendants act as a third-party "eavesdropper," redirecting Plaintiff's and ChatGPT API Class Members' electronic communications to Defendants' own servers for appropriation, and training of

their Products.

278. Defendants' interception of the contents of Plaintiff's and ChatGPT API Class Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Plaintiff's and ChatGPT API Class Members' friends, colleagues, or third-party entities. As described above, the ChatGPT API is designed to simultaneously intercept and send a recording of each keystroke, mouse click, movement, writing, or other data, activity, or intelligence to Defendants sufficient to not only identify Plaintiff and ChatGPT API Class Members but also to be able to understand, collect, and use for training Plaintiff's and ChatGPT API Class Members' communications.

279. **Unauthorized Purpose**. Plaintiff and ChatGPT API Class Members did not authorize Defendants to acquire, access, or intercept the content of their communications on third party platforms, websites, and applications. Therefore, such interception and recording of communications invades Plaintiff's and ChatGPT API Class Members' privacy. Defendants intentionally intercepted the contents of Plaintiff's and ChatGPT API Class Members' electronic communications for the purpose of committing a tortious act in violation of the Constitution or laws of the United States or of any State – namely, the knowing intrusion into a private place, conversation, or matter that would be highly offensive to a reasonable person.

280. **While in Transmission.** Through this calculated scheme of using ChatGPT API to intercept, acquire, transmit, and record Plaintiff's and ChatGPT API Class Members' electronic communications, Defendants willfully and without valid consent from all parties to the communication, take unauthorized measures to read and understand the contents or meaning of the electronic communications of Plaintiff and ChatGPT API Class. The interception and recording of electronic communications occurs while the electronic communications are in transit or passing over any wire, line, or cable, or are being sent from or received at any place.

281. In sending and in acquiring the content of Plaintiff's and ChatGPT API Class Members' communications with third-party platforms, Defendants' purpose was tortious, and designed to violate federal and state legal laws. By intentionally using, or endeavoring to use, the contents of the electronic communications of Plaintiff and ChatGPT API Class and Subclass

Members, while knowing or having reason to know that the information was obtained through the interception of an electronic communication, Defendants violate 18 U.S.C. § 2511(1)(a).

282. Plaintiff, individually and on behalf of the GPT API Class and Subclass Members, seeks all monetary and non-monetary relief allowed by law, including actual damages, statutory damages, punitive damages, preliminary and other equitable or declaratory relief, and attorneys' fees and costs.

## II. Microsoft's Interception of Communications Between ChatGPT Class Members

283. The allegations for violation of 18 U.S.C. § 2510 arise out of Defendant Microsoft's interception of Plaintiff's and ChatGPT User Class Members' communications which occurred on ChatGPT platform.

284. The transmissions of Plaintiff's and ChatGPT User Class Members' communications (including but not limited to chats, comments, replies, searches, keystrokes, mouse clicks/movements, signals, browser activity, or other data, activity, or intelligence) on ChatGPT platform qualify as "communications" under 18 U.S.C. § 2510(12).

285. By integrating ChatGPT technology on third party platforms, Defendants are in the unique position of having unrestricted, real-time access to the users' every input, move, mouse click, chat, comment, reply, search, keystroke, browser activity, or other data, activity, or intelligence on the third-party platform.

286. As Plaintiff and ChatGPT User Class Members interact with each other or the third-party entities, Defendant OpenAI intentionally divulges and Defendant Microsoft intentionally taps, electrically or otherwise intercepts, the lines of internet communications between Plaintiff, ChatGPT, and/or third party entities (integrated within ChatGPT through plug-in technologies).

287. In disregard for Plaintiff's and ChatGPT User Class Members' privacy rights, Defendant Microsoft acts as a third-party "eavesdropper," redirecting Plaintiff's and ChatGPT User Class Members' electronic communications to Defendant Microsoft's own servers for appropriation, and training of their Products.

288. Defendant Microsoft's interception of the contents of Plaintiff's and ChatGPT User Class Members' communications happens contemporaneously with their exchange of such

communications, whether such communications are directed to Defendant OpenAI or third-party entities. As described above, ChatGPT is designed to simultaneously intercept and send a recording of each keystroke, mouse click, movement, writing, or other data, activity, or intelligence to Defendant Microsoft sufficient to not only identify Plaintiff and ChatGPT User Class Members, but also to be able to understand, collect, and use for training Plaintiff's and ChatGPT User Class Members' communications.

289. **Unauthorized Purpose**. Plaintiff and ChatGPT User Class Members did not authorize Defendant Microsoft to acquire, access, or intercept the content of their communications on third party platforms, websites, and applications. Moreover, Plaintiff and ChatGPT User Class Members did not authorize either Defendant to train their AI Products on private information acquired by Defendants. Therefore, such interception and recording of communications invades Plaintiff's and ChatGPT User Class Members' privacy. Defendant OpenAI illegally divulged the content of such communications to Defendant Microsoft. Defendant Microsoft intentionally intercepted the contents of Plaintiff's and ChatGPT User Class Members' communications for the purpose of committing a tortious act in violation of the Constitution or laws of the United States or of any State – namely, the knowing intrusion into a private place, conversation, or matter that would be highly offensive to a reasonable person.

290. **While in Transmission.** Through this calculated scheme of using ChatGPT technology to intercept, acquire, transmit, and record Plaintiff's and ChatGPT User Class Members' electronic communications, Defendant Microsoft willfully and without any iota of valid consent from all parties to the communications, takes unauthorized measures to read and understand the contents or meaning of the electronic communications of Plaintiff and ChatGPT User Class Members. The interception and recording of electronic communications occur while the electronic communications are in transit or passing over any wire, line, or cable, or are being sent from or received at any place.

291. In sending and in acquiring the content of Plaintiff's and Class Members' communications with third-party platforms, Defendants' purpose was tortious, and designed to violate federal and state laws. By intentionally using, or endeavoring to use, the contents of the

electronic communications of Plaintiff and ChatGPT User Class Members, while knowing or having reason to know that the information was obtained through the interception of an electronic communication, Defendant Microsoft violates 18 U.S.C. § 2511(1)(a).

292. Plaintiff, individually and on behalf of the ChatGPT User Class Members, seeks all monetary and non-monetary relief allowed by law, including actual damages, statutory damages, punitive damages, preliminary and other equitable or declaratory relief, and attorneys' fees and costs.

## III. Defendant OpenAI's Interception of Microsoft User Class Members occurred on Microsoft's Websites, Platforms, Applications, and Programs which have integrated ChatGPT.

293. The allegations for violation of 18 U.S.C. § 2510 arise out of Defendant OpenAI's interception of Microsoft User Class Members' (collectively "Microsoft Subclasses") communications with their friends, family, colleagues, or other individuals or third-party entities, which occurred on Microsoft platforms (Teams, Bing, Outlook etc.), which integrate ChatGPT API.

294. The transmissions of Plaintiff's and Microsoft Subclasses' communications (including but not limited to chats, comments, replies, searches, keystrokes, signals, mouse clicks/movements, signals, browser activity, or other data, activity, or intelligence) on Microsoft's various applications, programs, platforms, websites which integrate ChatGPT API qualify as "communications" under 18 U.S.C. § 2510(12).

295. By integrating ChatGPT technology within the entire Microsoft suite, Defendant OpenAI is in the unique position of having unrestricted, real-time access to the users' every input, move, mouse click, chat, comment, reply, search, keystroke, browser activity, or other data, activity, or intelligence.

296. As Plaintiff and Microsoft Subclasses interact with each other or the third-party entities, Defendants intentionally tap, electrically or otherwise intercept, the lines of internet communications between Plaintiff, Microsoft Subclasses, and/or third-party entities.

297. In disregard for Plaintiff's and Microsoft Subclasses Members' privacy rights, Defendant OpenAI acts as a third-party "eavesdropper," redirecting Plaintiff's and Microsoft Subclasses Members' electronic communications to Defendants' own servers for appropriation, and

training of their Products.

298. Defendant OpenAI's interception of the contents of Plaintiff's and Microsoft Subclasses Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Plaintiff's and Microsoft Subclasses Members' friends, colleagues, or third-party entities. As described above, the ChatGPT API is designed to simultaneously intercept and send a recording of each keystroke, mouse click, signal, movement, writing, or other data, activity, or intelligence to Defendants sufficient to not only identify Plaintiff and Microsoft Subclasses Members, but also to be able to understand, collect, and use for training Plaintiff's and Microsoft Subclasses Members' communications.

299. **Unauthorized Purpose**. Plaintiff and Microsoft Subclasses did not authorize Defendant OpenAI to acquire, access, or intercept the content of their communications which occurred on Microsoft platforms, applications, programs, and websites. Therefore, such interception and recording of communications invades Plaintiff's and Microsoft Subclasses Members' privacy. Defendant OpenAI intentionally intercepted (and continues to intercept) the contents of Plaintiff's and Microsoft Subclasses Members' electronic communications for the purpose of committing a tortious act in violation of the Constitution or laws of the United States or of any State – namely, the knowing intrusion into a private place, conversation, or matter that would be highly offensive to a reasonable person.

300. **While in Transmission.** Through this calculated scheme of using ChatGPT API to intercept, acquire, transmit, and record Plaintiff's and Microsoft Subclasses Members' electronic communications, Defendant OpenAI willfully and without any iota of valid consent from all parties to the communication, implements unauthorized measures to read and understand the contents or meaning of Plaintiff's and Microsoft Subclasses' communications. The interception and recording of electronic communications occur while the electronic communications are in transit or passing over any wire, line, or cable, or are being sent from or received at any place.

301. In sending and in acquiring the content of Plaintiff's and Class Members' communications with third-party platforms, Defendant OpenAI's purpose was tortious, and designed to violate federal and state laws. By intentionally using, or endeavoring to use, the contents

of Plaintiff's and Microsoft Subclasses' electronic communications, while knowing or having reason to know that the information was obtained through the interception of an electronic communication, Defendant OpenAI violated and continues to violate 18 U.S.C. § 2511(1)(a).

302. Plaintiff, individually and on behalf of the Microsoft Subclasses Members, seeks all monetary and non-monetary relief allowed by law, including actual damages, statutory damages, punitive damages, preliminary and other equitable or declaratory relief, and attorneys' fees and costs.

**COUNT TWO: VIOLATION OF THE COMPUTER FRAUD AND ABUSE ACT, 18 U.S.C. § 1030**
**(on behalf of All Plaintiffs against Defendants)**

303. Plaintiff hereby incorporate Paragraphs 1 through 302 as if fully stated herein.

304. Plaintiff's, the Classes', and Subclasses' computer devices (including but not limited to Mac and Windows devices) were used for interstate communication and commerce and are therefore "protected computers" under 18 U.S.C. § 1030(e)(2)(B).

305. Defendants intentionally accessed Plaintiff's and the Classes and Subclasses Members' protected computers and obtained information thereby, and in doing so exceeded authority granted by Plaintiff, the Classes, and Subclasses to access the protected computers in violation of 18 U.S.C. § 1030(a)(2)(C). Plaintiff, the Classes, and Subclasses Members have a civil cause of action for violation of the CFAA under 18 U.S.C. § 1030(g) and have suffered damage or loss.

306. **Chat GPT Plug-In:** Defendants owned and operated their Products and ChatGPT Plug-Ins. Defendants integrated ChatGPT Plug-Ins within various platforms, websites, applications, and programs, and thereby intercepted and obtained Plaintiff's, the Classes', and Subclasses' Private Information, inclusive of keywords, mouse clicks, searches, movements, signals, and other activity and intelligence.

307. **Microsoft GPT Plug-In:** Defendant Microsoft owned and operated its Microsoft platforms, websites, programs, and applications which integrated Defendants' ChatGPT Plug-In. Defendant OpenAI intercepted and obtained Plaintiff's, the Classes', and Subclasses' Private Information, inclusive of keywords, mouse clicks, searches, movements, signals, and other activity

and intelligence. Defendants collected and transmitted this data to their Products, and used it to train their Products. Defendants' collected data allows Defendant to determine individual users' precise locations, unique identifiers, cookies, patterns (including browsing patterns, conversational patterns), conversational and browsing activities and habits, and a plethora of other Private Information.

308. **ChatGPT:** Defendant OpenAI owned and operated its ChatGPT platforms. Defendant OpenAI transmits all data from its ChatGPT platforms to Defendant Microsoft; Defendant Microsoft thereby intercepted and obtained Plaintiff's, the Classes', and Subclasses' Private Information, inclusive of keywords, mouse clicks, searches, movements, signals, and other activity and intelligence. Defendants collected, and transmitted this data to their Products, and used it to train their Products. Defendants' collected data allows Defendant to determine individual users' precise locations, unique identifiers, cookies, patterns (including browsing patterns, conversational patterns), conversational and browsing activities and habits, and a plethora of other Private Information.

309. Defendants accessed and otherwise transmitted this data without authorized consent from Plaintiff, Classes, and Subclasses; or at a minimum, as discussed above, exceeded any consent that was given.

310. Defendants were actively involved in implementing the unlawful interception alleged herein and promoted the use of their Products to U.S. residents and other companies, knowing about the privacy violations alleged herein. Defendants are also liable for this conduct because it occurred pursuant to the common enterprise of which they are a part.

311. Defendants' conduct caused "loss to 1 or more persons during any 1-year period . . . aggregating at least $5,000 in value" under 18 U.S.C. § 1030(c)(4)(A)(i)(I) because the unauthorized access and collection of Private Information (i) caused a diminution in value of Plaintiff's, Classes', and Subclasses' Private information, both of which occurred to millions of individuals, easily aggregating at least $5,000 in value.

312. For these reasons, and those discussed in this Complaint, Plaintiff, Classes, and Subclasses are entitled to "maintain a civil action against the violator to obtain compensatory

damages and injunctive relief or other equitable relief." 18 U.S.C. § 1030(g).

**COUNT THREE: VIOLATION OF THE CALIFORNIA INVASION OF PRIVACY ACT ("CIPA"), CAL. PENAL CODE § 631, *et seq.***
**(on behalf of Plaintiff and the ChatGPT, ChatGPT API User, and Microsoft User Classes against Defendants)**

313. Plaintiff hereby incorporates Paragraphs 1 through 253 as if fully stated herein.

314. The California Invasion of Privacy Act ("CIPA") is codified at Cal. Penal Code §§ 630 to 638. The Act begins with its statement of purpose:

> The Legislature hereby declares that advances in science and technology have led to the development of new devices and techniques for the purpose of eavesdropping upon private communications and that the invasion of privacy resulting from the continual and increasing use of such devices and techniques has created a serious threat to the free exercise of personal liberties and cannot be tolerated in a free and civilized society.

Cal. Penal Code § 630.

315. California Penal Code § 631(a) provides, in pertinent part:

> Any person who, by means of any machine, instrument, or contrivance, or in any other manner . . . willfully and without the consent of all parties to the communication, or in any unauthorized manner, reads, or attempts to read, or to learn the contents or meaning of any message, report, or communication while the same is in transit or passing over any wire, line, or cable, or is being sent from, or received at any place within this state; or who uses, or attempts to use, in any manner, or for any purpose, or to communicate in any way, any information so obtained, or who aids, agrees with, employs, or conspires with any person or persons to lawfully do, or permit, or cause to be done any of the acts or things mentioned above in this section, is punishable by a fine not exceeding two thousand five hundred dollars . . . .

316. California Penal Code § 632(a) provides, in pertinent part:

> A person who, intentionally and without the consent of all parties to a confidential communication, uses an electronic amplifying or recording device to eavesdrop upon or record the confidential communication, whether the communication is carried on among the parties in the presence of one another or by means of a telegraph, telephone, or other device, except a radio, shall be punished by a fine not exceeding two thousand five hundred dollars . . . .

317. Under either section of the CIPA, a defendant must show it had the consent of <u>all</u>

parties to a communication.

318. OpenAI has its principal place of business in California; designed, contrived, and effectuated its scheme to track users from California; and has adopted California substantive law to govern its relationship with its users. Defendants conspired with OpenAI to effectuate these schemes in and through California.

319. At all relevant times, Defendants' tracking and interceptions of Plaintiff's and Class members' internet communications was without authorization and consent from the Plaintiff, Class members, and the websites they were browsing. The interception by Defendants was unlawful and tortious.

320. Defendants' non-consensual tracking of Plaintiff's and Class members' internet communications was designed to attempt to learn at least some meaning of the content in the URLs and the communications that Plaintiff and Class members were engaged in.

321. The following items constitute "machine[s], instrument[s], or contrivance[s]" under the CIPA, and even if they do not, Defendants' deliberate and admittedly purposeful scheme that facilitated its interceptions falls under the broad statutory catch-all category of "any other manner":

a. The computer codes and programs Defendants used to track Plaintiff's and Class members' communications;

b. Plaintiff's and Class members' browsers and mobile applications;

c. Plaintiff's and Class members' computing and mobile devices;

d. Defendants' web and ad servers;

e. The web and ad-servers of websites from which Defendants tracked and intercepted Plaintiff's and Class members' communications;

f. The computer codes and programs that Defendants used to effectuate tracking and interception of Plaintiff's and Class members' communications; and

g. The plan Defendants carried out to effectuate the tracking and interception of Plaintiff's and Class members' communications.

322. The data collected by Defendants constituted "confidential communications," as that term is used in Section 632, because Plaintiff and Class members had objectively reasonable

expectations of privacy that the information would not be used for Defendants' AI products.

323. Plaintiff and Class members have suffered loss by reason of these violations, including, but not limited to, violation of their rights to privacy and loss of value in their personally-identifiable information.

324. Pursuant to California Penal Code § 637.2, Plaintiff and Class members have been injured by the violations of California Penal Code §§ 631 and 632, and each seek damages for the greater of $5,000 or three times the amount of actual damages, as well as injunctive relief

325. Plaintiff brings this claim individually and on behalf of the members of the proposed Classes against Defendants.

326. CIPA § 631(a) imposes liability for "distinct and mutually independent patterns of conduct." *Tavernetti v. Superior Ct.*, 22 Cal. 3d 187, 192 (1978). Thus, to establish liability under CIPA § 631(a), a plaintiff need only establish that the defendant, "by means of any machine, instrument, contrivance, or in any other manner," does any of the following:

> Intentionally taps, or makes any unauthorized connection, whether physically, electrically, acoustically, inductively or otherwise, with any telegraph or telephone wire, line, cable, or instrument, including the wire, line, cable, or instrument of any internal telephonic communication system,
>
> **OR**
>
> Willfully and without the consent of all parties to the communication, or in any unauthorized manner, reads or attempts to read or learn the contents or meaning of any message, report, or communication while the same is in transit or passing over any wire, line or cable or is being sent from or received at any place within this state,
>
> **OR**
>
> Uses, or attempts to use, in any manner, or for any purpose, or to communicate in any way, any information so obtained,
>
> **OR**
>
> Aids, agrees with, employs, or conspires with any person or persons to unlawfully do, or permit, or cause to be done any of the acts or things mentioned above in this section.

Cal. Penal Code § 631 (Deering 2023).

327. Section 631(a) is not limited to phone lines, but also applies to "new technologies" such as computers, the Internet, and email. *See Matera v. Google Inc.*, No. 15-CV-04062-LHK, 2016 U.S. Dist. LEXIS 107918, at *61-*63 (N.D. Cal. Aug. 12, 2016) (CIPA applies to "new

technologies" and must be construed broadly to effectuate its remedial purpose of protecting privacy); *Bradley v. Google, Inc.*, 2006 WL 3798134, at *5-6 (N.D. Cal. Dec. 22, 2006) (CIPA governs "electronic communications"); *In re Facebook, Inc. Internet Tracking Litigation*, 956 F.3d 589, 598-99 (9th Cir. 2020) (reversing dismissal of CIPA and common law privacy claims based on Facebook's collection of consumers' Internet browsing history).

328. Defendants' ChatGPT platform is a "machine, instrument, contrivance, or ... other manner" used to engage in the prohibited conduct at issue here.

**I. Defendants' Interception of Communications of ChatGPT API Class Members occurred on Third-Party Websites, Platforms, Applications, Programs which have integrated ChatGPT API. [Microsoft User Subclass is Excluded]**

329. The allegations for violation of CIPA § 631(a) arise out of Defendants' interception of Plaintiff's and ChatGPT API Class Members' (collectively referred to as ChatGPT API Class and Subclass) communications which occurred on various applications, platforms, and websites which integrate ChatGPT technology (*i.e.*, Stripe, Snapchat, etc.).

330. The transmissions of Plaintiff's and ChatGPT API Class Members' communications (including but not limited to chats, comments, replies, searches, keystrokes, mouse clicks/movements, signals, browser activity, or other data, activity, or intelligence) on various applications, programs, platforms, and websites which integrate ChatGPT API (*i.e.*, Stripe, Snapchat, etc.) qualify as "electronic communications" under Cal. Penal Code §629.51(2).

331. By incorporating ChatGPT technology on third party platforms, Defendants are in the unique position of having unrestricted, real-time access to the users' every input, move, chat, comment, reply, search, keystroke, or other browser activity/communication on the third-party platform.

332. As Plaintiff and ChatGPT API Class Members interact with the third-party platform, Defendants intentionally tap, electrically or otherwise, the lines of internet communication between Plaintiff and ChatGPT API Class Members, and/or third-party entities.

333. In disregard for Plaintiff's and ChatGPT API Class Members' privacy rights, Defendants act as a third-party "eavesdropper", redirecting Plaintiff's and ChatGPT API Members' electronic communications to Defendants' own servers for appropriation, and training of their

Products.

334. Defendants' interception of the contents of Plaintiff's and ChatGPT API Class Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Plaintiff's and ChatGPT API Class Members' friends, colleagues, or third-party entities. As described above, the ChatGPT technology, integrated on various platforms, is designed to simultaneously intercept and send a recording of each keystroke, mouse click, movement, writing, or other data, activity, or intelligence to Defendants sufficient to not only identify Plaintiff and ChatGPT API Class Members, but also to be able to understand, collect, and use for training Plaintiff's and ChatGPT API Class Members' communications.

335. Through this calculated scheme of using ChatGPT technology, integrated on various non-ChatGPT platforms (such as Snapchat, Stripe etc.) to intercept, acquire, transmit, and record Plaintiff's and ChatGPT API Class Members' electronic communications, Defendants willfully and without valid consent from all parties to the communications, take unauthorized measures to read and understand the contents or meaning of the electronic communications of Plaintiff and ChatGPT API Class Members. The interception and recording of electronic communications occurs while the electronic communications are in transit or passing over any wire, line, or cable, or are being sent from or received at any place.

336. Plaintiff and ChatGPT API Class Members did not authorize Defendants to acquire the content of their communications for the purposes of training Defendants' Products.

337. Plaintiff, individually and on behalf of the GPT API Class, also seeks all monetary and non-monetary relief allowed by law, including actual damages, statutory damages in accordance with § 637.2(a), punitive damages, preliminary and other equitable or declaratory relief, and attorneys' fees and costs.

## II. Microsoft's Interception of ChatGPT User Class Members' Communications on ChatGPT

338. The allegations for violation of CIPA § 631(a) arise out of Defendant Microsoft's interception of Plaintiff's and ChatGPT User Class Members' communications which occurred on ChatGPT platform.

339. The transmissions of Plaintiff's and ChatGPT User Class Members' communications (including but not limited to chats, comments, replies, searches, keystrokes, mouse clicks/movements, signals, browser activity, or other data, activity, or intelligence) on ChatGPT qualify as "electronic communications" under Cal. Penal Code § 629.51(2).

340. By developing ChatGPT and controlling the extent of training/development of this program, Defendants are in the unique position of having unrestricted, real-time access to the users' every input, move, mouse click, chat, comment, reply, search, keystroke, browser activity, or other data, activity, or intelligence on ChatGPT.

341. As Plaintiff and ChatGPT User Class Members ask questions, or otherwise interact with Defendant OpenAI, Defendant OpenAI intentionally aids and abets Defendant Microsoft to intentionally tap and intercept, electrically or otherwise, the lines of internet communications of Plaintiff's and ChatGPT User Class Members' searches and communications.

342. In disregard for Plaintiff's and ChatGPT User Class Members' privacy rights, Defendant Microsoft acts as a third-party "eavesdropper," redirecting Plaintiff and ChatGPT User Class Members' electronic communications to Defendant Microsoft's own servers for appropriation, and training of their Products.

343. Defendant Microsoft's interception of the contents of Plaintiff's and ChatGPT User Class Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Defendant OpenAI or third-party entities (for instance, Expedia). As described above, the ChatGPT technology is designed to simultaneously intercept and send a recording of each keystroke, mouse click, movement, writing, or other data, activity, or intelligence to Defendant Microsoft sufficient to not only identify Plaintiff and ChatGPT User Members, but also to be able to understand, collect, and use for training Plaintiff's and ChatGPT User Class Members' communications.

344. Defendant Microsoft intercepted communications including all text entry input as a search within ChatGPT as well as intercepted numerous other forms of a user's navigation and interaction with ChatGPT.

345. Through this calculated scheme of using ChatGPT to intercept, acquire, transmit, and

record Plaintiff's and ChatGPT User Class Members' electronic communications, Defendant Microsoft willfully and without any iota of valid consent from all parties to the communication, takes unauthorized measures to read and understand the contents or meaning of the electronic communications of Plaintiff and ChatGPT User Class Members. The interception and recording of electronic communications occur while the electronic communications are in transit or passing over any wire, line, or cable, or are being sent from or received at any place.

346. In sending and in acquiring the content of Plaintiff's and Class Members' communications on ChatGPT, Defendants' purpose was tortious, and designed to violate federal and state laws. By intentionally using, or endeavoring to use, the contents of the electronic communications of Plaintiff and ChatGPT User Class Members, while knowing or having reason to know that the information was obtained through the interception of an electronic communication, Defendant Microsoft violates CIPA § 631(a).

347. Additionally, under the fourth clause of § 631(a), Defendant OpenAI aided, agreed with, and conspired with Defendant Microsoft to accomplish the wrongful conduct at issue here. *Graham v. Noom, Inc.*, 533 F. Supp. 3d 823, 831-32 (N.D. Cal. 2021) (while a party to a communication may record the communication without triggering § 631(a) liability, it will be subject to derivative liability where the third party is liable for recording the communications in violation of the first, second or third clauses of § 631(a)); *Revitch v. New Moosejaw, LLC*, No. 18-cv-06827-VC, 2019 WL 5485330, at *2 (N.D. Cal. 2019) (conversation participants may be liable because § 631 "was designed to protect a person placing or receiving a call from a situation where the person on the other end of the line permits an outsider to tap his telephone or listen in on the call.")

348. Plaintiff, individually and on behalf of the hatGPT User Class Members, seeks all monetary and non-monetary relief allowed by law, including actual damages, statutory damages, punitive damages, preliminary and other equitable or declaratory relief, and attorneys' fees and costs.

## III. Defendant OpenAI's Interception of Microsoft User Class Members occurred on Microsoft's Websites, Platforms, Applications, and Programs which have integrated ChatGPT.

349. The allegations for violation of CIPA § 631(a) arise out of Defendant OpenAI's interception of Microsoft User Class Members' (collectively "Microsoft Subclass") communications with their friends, family, colleagues, or other individuals or third-party entities, which occurred on Microsoft platforms (Teams, Bing, Outlook etc.), which integrate ChatGPT API.

350. The transmissions of Plaintiff's and Microsoft Subclasses' communications (including but not limited to chats, comments, replies, searches, keystrokes, signals, mouse clicks/movements, browser activity, or other data, activity, or intelligence) on Microsoft's various applications, programs, platforms, and websites which integrate ChatGPT API qualify as "electronic communications" under Cal. Penal Code §629.51(2).

351. By integrating ChatGPT technology within the entire Microsoft suite, Defendant OpenAI is in the unique position of having unrestricted, real-time access to the users' every input, move, mouse click, chat, comment, reply, search, keystroke, browser activity, or other data, activity, or intelligence.

352. As Plaintiff and Microsoft Subclasses interact with each other or the third-party entities, Defendant OpenAI intentionally taps, electrically or otherwise intercepts, the lines of internet communications between Plaintiff, Microsoft Subclasses, and/or third-party entities.

353. In disregard for Plaintiff's and Microsoft Subclasses Members' privacy rights, Defendant OpenAI acts as a third-party "eavesdropper," redirecting Plaintiff's and Microsoft Subclasses Members' electronic communications to Defendants' own servers for appropriation, and training of their Products.

354. Defendant OpenAI's interception of the contents of Plaintiff's and Microsoft Subclasses Members' communications happens contemporaneously with their exchange of such communications on Microsoft platforms, whether such communications are directed to Plaintiff's and Microsoft Subclasses Members' friends, colleagues, or third-party entities. As described above, the ChatGPT API is designed to simultaneously intercept and send a recording of each keystroke, mouse click, signal, movement, writing, or other data, activity, or intelligence to Defendant OpenAI

sufficient to not only identify Plaintiff and Microsoft Subclasses Members, but also to be able to understand, collect, and use for training Plaintiff's and Microsoft Subclasses Members' communications.

355. Additionally, under the fourth clause of § 631(a), Defendant Microsoft aided, agreed with, and conspired with Defendant OpenAI to implement AI technology within its own platforms. The incorporation of such technology shares users' electronic communications with Microsoft platforms with OpenAI in an effort to accomplish the wrongful conduct at issue here. *Graham v. Noom, Inc.*, 533 F. Supp. 3d 823, 831-32 (N.D. Cal. 2021) (while a party to a communication may record the communication without triggering § 631(a) liability, it will be subject to derivative liability where the third party is liable for recording the communications in violation of the first, second or third clauses of § 631(a)); *Revitch v. New Moosejaw, LLC*, No. 18-cv-06827-VC, 2019 WL 5485330, at *2 (N.D. Cal. 2019) (conversation participants may be liable because § 631 "was designed to protect a person placing or receiving a call from a situation where the person on the other end of the line permits an outsider to tap his telephone or listen in on the call.")

356. Plaintiff, individually and on behalf of the Microsoft Subclasses Members, seeks all monetary and non-monetary relief allowed by law, including actual damages, statutory damages, punitive damages, preliminary and other equitable or declaratory relief, and attorneys' fees and costs.

357. Unless enjoined, Defendants will continue to commit the illegal acts alleged here.

358. Plaintiff and Class Members seek all relief available under Cal. Penal Code § 637.2, including injunctive relief and statutory damages of $5,000 per violation.

**COUNT FOUR: VIOLATION OF CALIFORNIA UNFAIR COMPETITION LAW**

**(Cal. Bus. & Prof. Code §§ 17200, *et seq.*)**
**(on behalf of Plaintiff and the Classes against Defendants)**

359. Plaintiff hereby incorporates Paragraphs 1 through 358 as if fully stated herein.

360. As discussed above, Plaintiff believes that California law should apply to all claimants, including out of state residents.

361. California Business & Professions Code, sections 17200, *et seq.* (the "UCL") prohibits unfair competition and provides, in pertinent part, that "unfair competition shall mean and

include unlawful, unfair or fraudulent business practices and unfair, deceptive, untrue or misleading advertising."

**I. Unlawful**

362. Defendants engaged in and continue to engage in "unlawful" business acts and practices under the Unfair Competition Law because Defendants took, accessed, intercepted, tracked, collected, or used Plaintiff's and Classes' Private Information, including but not limited to their private conversations, personally identifiable information, financial and medical data, keystrokes, searches, cookies, browser activity and other data, and shared this information with each other, while also using this information to train Defendants' AI Products. Defendants' unlawful conduct is as follows:

a) Web-Scraping and Interception of Communications, Private Information and Data: Defendants scraped nearly the entire internet in order to train their AI Products, and in this process, Defendants accessed, and stole private conversations, personal information, and other private data from websites including Reddit, Twitter, TikTok, Spotify, YouTube, and other websites, without consent of the individuals. Defendants' illegal web scraping violates privacy laws, and other laws outlined in this complaint. Defendants failed to register as data brokers under California law as required.

b) Defendants' Intercepted Communications and Accessed, Collected, and Tracked Private Information from Platforms Which Integrated ChatGPT: Defendants intercepted, tracked, and recorded communications, messages, chats, web activity, user activity, associated cookies, keystrokes and other Private Information through its ChatGPT technology integrated within hundreds of applications (including but not limited to Stripe, Snapchat, Expedia etc.) which were used to train Defendants' Products. Defendants' illegal tracking of such data, which is subsequently used to train Defendants' AI products violates privacy laws, California wiretapping law, and other laws outlined in this complaint.

c) OpenAI's Interception of Communications and Accessed, Collected, and Tracked

Private Information on Microsoft Platforms: Defendant Microsoft aided Defendant OpenAI in intercepting, tracking, and recording communications, messages, chats, web activity, user activity, associated cookies, and other Private Information through its ChatGPT technology integrated within the entire Microsoft suite (Microsoft Teams, Microsoft Outlook, Bing). Defendant's OpenAI illegal tracking of such data and Defendant Microsoft's aiding and abetting this conduct violates privacy laws, California wiretapping law, and other laws outlined in this complaint.

d) Microsoft's Interception of Communications and Accessed, Collected, and Tracked Private Information on ChatGPT: Defendant OpenAI aided Defendant Microsoft in intercepting, tracking, and recording communications, messages, chats, web activity, user activity, associated cookies, and other Private Information by sharing access to ChatGPT and sending all communications to Defendant Microsoft and its partners.

363. Defendants' conduct as alleged herein was unfair within the meaning of the UCL. The unfair prong of the UCL prohibits unfair business practices that either offend an established public policy or that are immoral, unethical, oppressive, unscrupulous, or substantially injurious to consumers.

364. Defendants' conduct violates the EPCA, CFAA, CIPA, California Consumer Privacy Act ("CCPA"), Cal. Civ. Code § 1798.100, *et seq.*, Section 5 of the Federal Trade Commission Act ("FTCA"), Cal. Bus. & Prof. Code § 22575, *et seq.,* and other tort claims stated in this lawsuit. The violations of EPCA, CFAA, CIPA, and other tort claims stated in this lawsuit, are incorporated herein by reference.

365. Under the CCPA, a business that collects consumers' personal information is required, at or before the point of collection, to provide notice to consumers indicating: (1) "[t]he categories of personal information to be collected and the purposes for which the categories of personal information are collected or used and whether that information is sold or shared"; (2) "the categories of sensitive personal information to be collected and the purposes for which the categories of sensitive personal information are collected or used, and whether that information is sold or shared."; and (3) "[t]he length of time the business intends to retain each category of personal

information . . .” Cal. Civ. Code § 1798.100(a).

366. “Personal information” is defined by the CCPA as “information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.” Cal. Civ. Code § 1798.140(v)(1).

367. As alleged, Defendants use web scraping technology to collect information from webpages across the internet and, in so doing, Defendants gather and compilepersonal information about consumers that is reflected on those webpages.

368. Because Defendants conduct web scraping across millions of web pages, without asking the affected consumers their permission to use their content for training, Defendants do not, and cannot provide consumers with the notice required by Cal. Civ. Code § 1798.100(a) at or before the point of collection. Similarly, Defendants intercept and wiretap users’ communications on various platforms which integrate ChatGPT, Microsoft platforms, and ChatGPT platforms, to use these intercepted communications and gathered data to train their Products. Defendants never notified Plaintiff and affected Class Members of this extensive wiretapping, and more importantly, that this information would be used for commercial purposes and development of Defendants’ Products. Therefore, Defendants failed to provide notice to the affected consumers as required by Cal. Civ. Code § 1798.100(a).

369. Defendant’s failure to provide notice to Plaintiff and Class Members whose personal information is collected through the process of web scraping and illegal wiretapping is unlawful and violates Cal. Civ. Code § 1798.100(a).

370. The CCPA further grants consumers the right to “request that a business that collects a consumer’s personal information disclose to that consumer the categories and specific pieces of personal information the business has collected.” Cal. Civ. Code § 1798.100(b).

371. Upon receipt of a verifiable request for disclosure pursuant to Section 1798.110, a business must “disclose any personal information it has collected about a consumer, directly or indirectly, including through or by a service provider or contractor, to the consumer . . .” Cal. Civ. Code § 1798.130 (3)(A).

372. Any disclosure must provide the requesting consumer with all of the following: (1)

"The categories of personal information it has collected about that consumer"; (2) "The categories of sources from which the personal information is collected"; (3) "The business or commercial purpose for collecting, selling, or sharing personal information"; (4) "The categories of third parties to whom the business discloses personal information"; and (5) "The specific pieces of personal information it has collected about that consumer." Cal. Civ. Code § 1798.110(a).

373. Consumers also "have the right to request that a business delete any personal information about the consumer which the business has collected from the consumer." Cal. Civ. Code § 1798.105(a).

374. Pursuant to Cal. Civ. Code §§ 1798.100(b) and 1798.130(a), OpenAI's privacy policy provides a method by which California residents who have had their data collected may request disclosure of the categories and specific pieces of personal information OpenAI has collected about them.[235] OpenAI's privacy policy specifically states that consumers "may have certain statutory rights in relation to their Personal Information," including the right to "Access your Personal Information."[236]

375. To exercise their right to access the Personal Information OpenAI has collected about them, consumers are instructed to email their request for disclosure to dsar@openai.com.[237]

376. Under the heading "Additional U.S. State Disclosures," the privacy policy states that some users may have "[t]he right to know information about our processing of your Personal Information, including the specific pieces of Personal Information that we have collected from you . . ."[238] Users are instructed that, "to the extent applicable under local law, [they] can exercise privacy rights. . . by submitting a request to dsar@openai.com."[239]

377. Yet OpenAI fails to disclose that once its AI Products have been trained on an individual's information, that information has been included into the product and cannot reasonably be extracted. Whether individuals' information was collected through web scraping or obtained

---

[235] *Privacy Policy*, OPENAI, https://openai.com/policies/privacy-policy (last updated November 14, 2023).
[236] *Id.*
[237] *Id.*
[238] *Id.*
[239] *Id.*

through interception from ChatGPT, or other platforms incorporating ChatGPT, this information, once used to train Products, cannot be extracted. Therefore, Defendants violated and continue to violate CCPA.

378. Plaintiff, individually and on behalf of the Classes, seeks: (i) an injunction requiring OpenAI to revise its privacy policy to fully disclose all information required under CCPA, and to delete all information previously collected in violation of these laws; (ii) relief under Cal. Bus. & Prof. Code § 17200, et seq., including, but not limited to, restitution to Plaintiff and other members of the Classes of money or property Defendants acquired by means of their unlawful business practices; and, as a result of bringing this action to vindicate and enforce an important right affecting the public interest, (iii) reasonable attorney's fees (pursuant to Cal. Code of Civ. P. § 1021.5).

379. Defendants' unlawful actions in violation of the UCL have caused and are likely to cause substantial injury to consumers that consumers cannot reasonably avoid themselves and that is not outweighed by countervailing benefits to consumers or competition.

380. As a direct and proximate result of Defendants' misconduct, Plaintiff and Class Members had their private communications containing information related to their sensitive and confidential Private Information intercepted, disclosed, and used by third parties, including but not limited to each Defendant.

381. As a result of Defendants' unlawful conduct, Plaintiff and Class Members suffered an injury, including violation to their rights of privacy, loss of value and privacy of their Private Information, loss of control over their sensitive personal information, and suffered embarrassment and emotional distress as a result of this unauthorized scraping, interception, sharing, and misuse of information.

**II. Unfair**

382. Defendants' conduct as alleged herein was unfair within the meaning of the UCL. The unfair prong of the UCL prohibits unfair business practices that either offend an established public policy or that are immoral, unethical, oppressive, unscrupulous or substantially injurious to consumers.

383. Defendants also engaged in business acts or practices deemed "unfair" under the UCL

because, as alleged above, Defendants failed to disclose that they scraped information belonging to millions of internet users without the users' consent. Defendants also failed to disclose that they used the stolen information to train their Products, without consent of the internet users. Furthermore, Defendants failed to disclose that they were intercepting, tracking Private Information belonging to millions of ChatGPT users, and the users of other platforms which integrated ChatGPT. Private Information obtained from individual uses of ChatGPT and other platforms which integrate ChatGPT was and is continuing to be used to train Defendants' Products, without consent of the users.

384. Unfair acts under the UCL have been interpreted using three different tests: (1) whether the public policy which is a predicate to a consumer unfair competition action under the unfair prong of the UCL is tethered to specific constitutional, statutory, or regulatory provisions; (2) whether the gravity of the harm to the consumer caused by the challenged business practice outweighs the utility of the defendant's conduct; and (3) whether the consumer injury is substantial, not outweighed by any countervailing benefits to consumers or competition, and is an injury that consumers themselves could not reasonably have avoided.

385. Under the UCL, a business practice that is likely to deceive an ordinary consumer constitutes a deceptive business practice. Defendants' conduct was deceptive in numerous respects.

386. Defendants' misrepresentations and omissions include both implicit and explicit representations.

387. Defendant OpenAI represented, throughout the Class Period, that it would "respect your privacy and [is] strongly committed to keeping secure any information we obtain from you or about you."

388. Defendants' conduct, as alleged herein, was fraudulent within the meaning of the UCL. Defendants made deceptive misrepresentations and omitted known material facts in connection with the solicitation, interception, disclosure, and use of Plaintiff's and Class Members' User Data. Defendants actively concealed and continued to assert misleading statements regarding their protection and limitation on the use of the User Data. Meanwhile, Defendants were collecting and sharing Plaintiff's and Class Members' User Data without their authorization or knowledge in

order to profit off of the information, and to deliver advertisements to Plaintiff and Class Members, among other unlawful purposes.

389. Defendants' conduct, as alleged herein, was unlawful within the meaning of the UCL because Defendants violated regulations and laws as discussed herein, including but not limited to HIPAA, Section 5 of the Federal Trade Commission Act ("FTCA"), 15 U.S.C. § 45 and the CIPA.

390. Defendants reaped profits from these actions in the form of increased company valuation, investments, improved language model performance, and dominance in the AI field.

391. Defendants' unlawful actions in violation of the UCL have caused and are likely to cause substantial injury to consumers that consumers cannot reasonably avoid themselves and that is not outweighed by countervailing benefits to consumers or competition.

392. As a direct and proximate result of Defendants' misconduct, Plaintiff and Class Members had their private communications containing information related to their sensitive and confidential User Data intercepted, disclosed, and used by third parties, including but not limited to each Defendant.

393. As a result of Defendants' unlawful conduct, Plaintiff and Class Members suffered an injury, including violation to their rights of privacy, loss of the privacy of their PHI/PII, loss of control over their sensitive personal information, and suffered aggravation, inconvenience, and emotional distress.

394. Further, Defendants' conduct is immoral, unethical, oppressive, unscrupulous and substantially injurious to Plaintiff and Class Members, and there are no greater countervailing benefits to consumers or competition.

395. Plaintiff, as well as Class Members, were harmed by Defendants' violations of Cal. Bus. & Prof. Code §17200. Defendants' practices were a substantial factor and caused injury in fact and actual damages to Plaintiff and Class Members.

396. As a direct and proximate result of Defendants' deceptive acts and practices, Plaintiff and Class Members have suffered and will continue to suffer an ascertainable loss of money or property, real or personal, and monetary and non-monetary damages, as described above, including the loss or diminishment in value of their Private Information and the loss of the ability to control

the use of their Private Information, which allowed Defendants to profit at the expense of Plaintiff and Class Members.

397. Plaintiff's and Class Members' Personal Information has tangible value; it is now in the possession of Defendants, who have used and will continue to use it for financial gain.

398. Plaintiff's and Class Members' injury was the direct and proximate result of Defendants' conduct described herein.

399. Defendants' retention of Plaintiff's and Class Members' Personal Information presents a continuing risk to them as well as the general public.

400. Plaintiff, individually and on behalf of the Class Members, seek: (1) an injunction requiring Defendants to permanently delete, destroy or otherwise sequester the Private Information collected without consent; (2) compensatory restitution of Plaintiff's and Class Members' money and property lost as a result of Defendants' acts of unfair competition; (3) disgorgement of Defendants' unjust gains; and (4) reasonable attorney's fees (pursuant to Cal. Code of Civ. Proc. § 1021.5).

401. Had Plaintiff and Class Members known Defendants would disclose and misuse their User Data in contravention of Defendants' representations, they would not have used Defendants' Products.

402. Defendants' unlawful actions in violation of the UCL have caused and are likely to cause substantial injury to consumers that consumers cannot reasonably avoid themselves and that is not outweighed by countervailing benefits to consumers or competition.

403. As a direct and proximate result of Defendants' misconduct, Plaintiff and Class Members had their private communications containing information related to their sensitive and confidential Private Information intercepted, disclosed, and used by Defendants, to train their Products.

404. As a result of Defendants' unlawful conduct, Plaintiff and Class Members suffered an injury, including violation to their rights of privacy, loss of the privacy of their Private Information, and loss of control over their sensitive personal information, and suffered aggravation, inconvenience, and emotional distress.

**COUNT FIVE: NEGLIGENCE**
**(on behalf of Plaintiff and the Classes against Defendants)**

405. Plaintiff hereby incorporates Paragraphs 1 through 404 as if fully stated herein.

406. Defendants owed a duty to Plaintiff and the Classes to exercise due care in: (a) obtaining data to train their Products; (b) not using individual's private information to train Defendants' AI; (c) ensuring that individuals' private data is not shared with or disclosed to unauthorized parties (including Defendant Microsoft); (d) destroying personal information to which Defendants had no legal right to possess.

407. Defendants' duties to use reasonable care arose from several sources, including those described below. Defendants had a common law duty to prevent foreseeable harm to others, including Plaintiff and members of the Classes, who were the foreseeable and probable victims of Defendants' unlawful practices. Defendants acknowledge the Products are inherently unpredictable and may even evolve to act against human interests. Nevertheless, Defendants collected and continue to collect Private Information of millions of individuals and permanently feed the data to the Products, to train the Products for Defendants' commercial benefit. Defendants knowingly put Plaintiff and the Classes in a zone of risk that is incalculable – but unacceptable by any measure of responsible data protection and use.

408. Defendants' conduct as described above constituted an unlawful breach of their duty to exercise due care in collecting, storing, and safeguarding Plaintiff's and the Class Members' Private Information by failing to protect this information.

409. Plaintiff and Class Members trusted Defendants to act reasonably, as a reasonably prudent manufacturer of AI products, and also trusted Defendants not to use individuals' Private Information to train their AI products. Defendants failed to do so, and breached their duty.

410. Defendants' negligence was, at least, a substantial factor in causing Plaintiff's and the Classes' Private Information to be improperly accessed, disclosed, used for development and training of a dangerous product, and in causing the Class members' injuries.

411. The damages suffered by Plaintiff and the Class members was the direct and reasonably foreseeable result of Defendants' negligent breach of their duties to adequately design,

implement, and maintain reasonable practices to (a) avoid web scraping without consent of the users; (b) avoid using Personal Information to train their AI products; and (c) avoid collecting and sharing Users' data with each other.

450. Defendants' negligence directly caused significant harm to Plaintiff and the Classes.

**COUNT SIX: INVASION OF PRIVACY**
**(on behalf of Plaintiff and the Classes against Defendants)**

412. Plaintiff hereby incorporates Paragraphs 1 through 411 as if fully stated herein.

413. The right to privacy in California's Constitution creates a right of action against private entities such as Defendants.

414. Plaintiff's and Class members' expectation of privacy is deeply enshrined in California's Constitution. Article I, section 1 of the California Constitution provides: "All people are by nature free and independent and have inalienable rights. Among these are enjoying and defending life and liberty, acquiring, possessing, and protecting property and pursuing and obtaining safety, happiness, *and privacy*." (Emphasis added).

415. The phrase "and privacy" was added in 1972 after voters approved a proposed legislative constitutional amendment designated as Proposition 11. Critically, the argument in favor of Proposition 11 reveals that the legislative intent was to curb businesses' control over the unauthorized collections and use of consumers' personal information, stating:

> The right of privacy is the right to be left alone…It prevents government and business interests from collecting and stockpiling unnecessary information about us and from misusing information gathered for one purpose in order to serve other purposes or to embarrass us. Fundamental to our privacy is the ability to control circulation of personal information. This is essential to social relationships and personal freedom.

416. The principal purpose of this constitutional right was to protect against unnecessary information gathering, use, and dissemination by public and private entities, including Defendants.

417. To plead a California constitutional privacy claim, a plaintiff must show an invasion of: 1) a legally protected privacy interest; 2) where the plaintiff had a reasonable expectation of privacy in the circumstances; and 3) conduct by the defendant constituting a serious invasion of privacy.

418. As described herein, Defendants have intruded upon the following legally protected privacy interests:

a. The Federal Wiretap Act as alleged herein;

b. The California Wiretap Act as alleged herein;

c. A Fourth Amendment right to privacy contained on personal computing devices, including web-browsing activity, as explained by the United States Supreme Court in the unanimous decision of *Riley v. California*;

d. The California Constitution, which guarantees Californians the right to privacy; and

e. Defendant's Privacy Policies and policies referenced therein.

419. Plaintiff and Class members had a reasonable expectation of privacy under the circumstances in that Plaintiff and Class members could not reasonably expect Defendants would commit acts in violation of federal and state civil and criminal laws.

420. Defendants' actions constituted a serious invasion of privacy in that they:

a. Invaded a zone of privacy protected by the Fourth Amendment, namely the right to privacy in data contained on personal computing devices, including web search and browsing histories;

b. Violated several federal criminal laws, including the Wiretap Act;

c. Violated dozens of state criminal laws on wiretapping and invasion of privacy, including the California Invasion of Privacy Act;

d. Invaded the privacy rights of hundreds of millions of Americans (including Plaintiff and class members) without their consent;

e. Constituted the unauthorized taking of valuable information from hundreds of millions of Americans through deceit; and

f. Further violated Plaintiff's and Class members' reasonable expectation of privacy via Defendants' review, analysis, and subsequent uses of Plaintiff's and Class members' browsing activity that Plaintiff and Class members considered sensitive and confidential, and did not intend to be used in Defendants' AI products.

421. Committing criminal acts against hundreds of millions of Americans constitutes an egregious breach of social norms that is highly offensive.

422. The surreptitious and unauthorized tracking of the internet communications of millions of Americans constitutes an egregious breach of social norms that is highly offensive.

423. Defendants' intentional intrusion into Plaintiff's and Class members' internet communications and their computing devices and web-browsers was highly offensive to a reasonable person in that Defendants violated federal and state criminal and civil laws designed to protect individual privacy and against theft.

424. The taking of personally-identifiable information from hundreds of millions of Americans through deceit is highly offensive behavior.

425. Secret monitoring of web browsing is highly offensive behavior.

426. Following Defendants' unauthorized interception of the sensitive and valuable personal information, the subsequent analysis and use of that activity to develop and refine Defendants' AI products violated Plaintiff's and Class Members' reasonable expectations of privacy.

427. Wiretapping and surreptitious recording of communications is highly offensive behavior.

428. Defendants' lacked any legitimate business interest in tracking users, then using that information in AI products without their consent.

429. Plaintiff and Class members have been damaged by Defendants' invasion of their privacy and are entitled to just compensation and injunctive relief.

**COUNT SEVEN: INTRUSION UPON SECLUSION**
**(on behalf of Plaintiff and the Classes against Defendants)**

430. Plaintiff hereby incorporates Paragraphs 1 through 429 as if fully stated herein.

431. Plaintiff asserting a claim for intrusion upon seclusion must plead: 1) intrusion into a private place, conversation, or matter; and 2) in a manner highly offensive to a reasonable person.

432. In carrying out their scheme to track and intercept Plaintiff's and Class members' communications and other private data, Defendants intentionally intruded upon Plaintiff's and Class members' solitude or seclusion in that Defendants effectively placed themselves in the middle of

conversations to which they were not authorized parties.

433. Defendants' actions were not authorized by Plaintiff and Class members, the Websites with which they were communicating, or the devices that Plaintiff and Class members were using to facilitate those communications.

434. Defendants' intentional intrusion into those communications and Plaintiff's and Class members' devices was highly offensive to a reasonable person in that they violated federal and state criminal and civil laws designed to protect individual privacy and against theft.

435. The taking of personally identifiable information from the hundreds of millions of Americans through deceit is highly offensive behavior.

436. Defendants' secret monitoring of web browsing is also highly offensive behavior.

437. Wiretapping and surreptitious recording of communications is highly offensive behavior.

438. Public polling on internet tracking has consistently revealed that the overwhelming majority of Americans believe it is important—or very important—to be "in control of who can get information" about them; to not be tracked without their consent; and to be in "control[] of what information is collected about [them]." This desire to control one's information is especially heightened in today's electronic age and with the proliferation of AI products.

439. Plaintiff and Class members have been damaged by Defendants' invasion of their privacy and are entitled to reasonable compensation including but not limited to disgorgement of profits related to the unlawful internet tracking, collection, and use of their data in AI products.

**COUNT EIGHT: LARCENY/RECEIPT OF STOLEN PROPERTY CAL. PENAL CODE §§ 496(a) and 496(c)**
**(on behalf of Plaintiff and the Classes against Defendants)**

440. Plaintiff hereby incorporates Paragraphs 1 through 439 as if fully stated herein.

441. Courts recognize that internet users have a property interest in their personal information and data. *See Calhoun v. Google, LLC,* 526 F. Supp. 3d 605, 635 (N.D. Cal. 2021) (recognizing property interest in personal information and rejecting Google's argument that "the personal information that Google allegedly stole is not property"); *In re Experian Data Breach Litigation,* SACV 15-1592 AG (DFMx), 2016 U.S. Dist. LEXIS 184500, at *14 (C.D. Cal. Dec. 29,

2016) (loss of value of PII is a viable damages theory); *In re Marriott Int'l Inc. Customer Data Sec. Breach Litig.,* 440 F. Supp. 3d 447, 460-61 (D. Md. 2020) ("The growing trend across courts that have considered this issue is to recognize the lost property value of this [personal] information."); *Simona Opris v. Sincera,* No. 21-3072, 2022 U.S. Dist. LEXIS 94192, at *20 (E.D. Pa. May 23, 2022) (collecting cases).

442. Defendants owned and operated their AI Products and GPT Platforms (ChatGPT, ChatGPT Plug-Ins, ChatGPT API). Defendants illegally obtained vast amounts of private information to train their AI Products.

**A. Defendants' Taking of Individual's Private Information to Train Their AI Violated Plaintiff's Property Interests**

443. Penal Code § 496(a) creates an action against "any" person who (1) receives "any" property that has been stolen or obtained in any manner constituting theft, knowing the property to be stolen or obtained, or (2) conceals, sells, withholds, or aids in concealing or withholding "any" property from the owner, knowing the property to be so stolen or illegally obtained.

444. Under Penal Code § 1.07(a)(38), "person" means "an individual, corporation, or association." Thus, Defendants are persons under section 496(a).

445. As discussed above, Defendants stole the contents of the internet – everything individuals posted, information about the individuals, personal data, medical information, and other information – all used to create their Products to generate massive profits. At no point did Defendants have individuals consent to take/scrape this information in order to train their AI Products. Defendants meet the grounds for liability under Cal. Penal Code 496(a) because each of them:

a. Knew that the taken information was stolen or obtained by theft, and with such knowledge;

b. Concealed, withheld, or aided in concealing or withholding said data from their rightful owners by unlawfully using the data to train their Products;

c. Defendants moved the data from the internet in order to feed it into their Products for training.

446. Pursuant to California Penal Code § 496(c), Plaintiff, individually and on behalf of

the Classes, seeks actual damages, treble damages, costs of suit, and reasonable attorneys' fees.

**B. Tracking, Collecting, and Sharing Private Information Without Consent**

447. As described above, in violation of Cal. Penal Code § 496(a) and (c), Defendants unlawfully collected, used, and exercised dominion and control of Private Information belonging to Plaintiff and Class Members.

448. Defendants wrongfully took Plaintiff's, ChatGpt User Class', ChatGPT API User Class', and Microsoft User Class' (collectively "User Classes") Private Information to be used to feed into Defendants' AI Products, to train and develop a dangerous technology.

449. Plaintiff and the User Classes Members did not consent to such taking and misuse of their personal data, and Private Information.

450. Defendants did not have consent from any state or local government agency allowing them to engage in such taking and misuse of Private Information.

451. Defendants' taking of Private Information was intended to deprive the owners of such information from the ability to use their Private Information in the way they chose.

452. Defendants did so to maximize their profits and become rich at the expense of Plaintiff and the Classes.

453. Defendants' collected data allows Defendants and their AI to learn the unique patterns of individuals, their online activities, habits, and speech/writing patterns.

454. Defendants moved Private Information to store and collect it on Defendant Microsoft's servers, and thereafter, fed it to their AI products.

455. As a result of Defendants' actions, Plaintiff and User Classes Members seek injunctive relief, in the form of Defendants' cessation of tracking practices in violation of state law, and destruction of all personal data obtained in violation of state law.

456. As a result of Defendants' actions, Plaintiff and the Classes seek nominal, actual, treble, and punitive damages in an amount to be determined at trial. Plaintiff and the Classes seek treble and punitive damages because Defendants' actions—which were malicious, oppressive, and willful—were calculated to injure Plaintiff and made in conscious disregard of Plaintiff's rights. Punitive damages are warranted to deter Defendants from engaging in future misconduct.

457. Plaintiff seeks restitution for the unjust enrichment obtained by Defendants as a result of the commercialization of Plaintiff's and the Classes' sensitive data.

**COUNT NINE: CONVERSION**
**(on behalf of Plaintiff and the Classes against Defendants)**

458. Plaintiff hereby incorporates Paragraphs 1 through 457 as if fully stated herein.

459. The Classes repeat and incorporate by reference all preceding paragraphs as if fully set forth herein.

460. Property is the right of any person to possess, use, enjoy, or dispose of a thing, including intangible things such as data or communications. Plaintiff's and Class Members' personal information is their property. *Calhoun v. Google* LLC, 526 F. Supp. 3d 605, 636 (N.D. Cal. 2021).

461. As described in the cause of action for Larceny/Receipt of Stolen Property, Cal. Penal Code § 496(a) and (c), Defendants unlawfully collected, used, and exercised dominion and control over the Class Members' personal and private information without authorization.

462. Defendants wrongfully exercised control over Plaintiff's and the Classes' information and have not returned it.

463. Plaintiff and the Class Members have been damaged as a result of Defendants' unlawful conversion of their property.

**COUNT TEN: UNJUST ENRICHMENT**
**(on behalf of Plaintiff and the Classes against Defendants)**

464. Plaintiff hereby incorporates Paragraphs 1 through 463 as if fully stated herein.

465. By virtue of the unlawful, unfair and deceptive conduct alleged herein, Defendants knowingly realized hundreds of millions of dollars in revenue from the use of the Personal Information of Plaintiff and Class Members for the commercial training of its ChatGPT and other AI language models.

466. This Private and Personal Information, the value of the Private and Personal Information, and/or the attendant revenue, were monetary benefits conferred upon Defendants by Plaintiff and the members of the Classes.

467. As a result of Defendants' conduct, Plaintiff and Class Members suffered actual

damages in the loss of value of their Private Information and the lost profits from the use of their Private Information.

468. It would be inequitable and unjust to permit Defendants to retain the enormous economic benefits (financial and otherwise) it has obtained from and/or at the expense of Plaintiff and Class Members.

469. Defendants will be unjustly enriched if they are permitted to retain the economic benefits conferred upon them by Plaintiff and Class Members through Defendants' obtaining the Private Information and the value thereof, and profiting from the unlawful, unauthorized, and impermissible use of the Private Information of Plaintiff and Class members.

470. Plaintiff and Class members are therefore entitled to recover the amounts realized by Defendants at the expense of Plaintiff and Class Members.

471. Plaintiff and the Classes have no adequate remedy at law.

472. Plaintiff and the members of the Classes are entitled to restitution, disgorgement, and/or the imposition of a constructive trust to recover the amount of Defendants' ill-gotten gains, and/or other sums as may be just and equitable.

**PRAYER FOR RELIEF**

WHEREFORE, Plaintiff, individually and on behalf of the Proposed Classes, respectfully requests the following relief:

A. Certify this action as a class action pursuant to Rule 23 of the Federal Rules of Civil Procedure;

B. Appoint Plaintiff to represent the Classes;

C. Appoint undersigned counsel to represent the Classes;

D. Award compensatory damages (including treble damages, where appropriate) to Plaintiff and the Class members against Defendants for all damages sustained as a result of Defendants' wrongdoing, in an amount to be proven at trial, including interest thereon;

E. Award statutory (including treble damages, where appropriate) damages to Plaintiff and the Class members against Defendants;

F. Award nominal damages to Plaintiff and the Class members against

Defendants;

G. Non-restitutionary disgorgement of all profits that were derived, in whole or in part, from Defendants' conduct;

H. Award punitive damages to Plaintiff and the Class members against Defendants;

I. For all Counts, permanently restrain Defendants, and its officers, agents, servants, employees, and attorneys, from the conduct at issue in this Action and otherwise violating its policies with consumers, and award all other appropriate injunctive and equitable relief deemed just and proper, including:

1. Establishment of an independent body of thought leaders (the "AI Council") who shall be responsible for approving uses of the Products before, not after, the Products are deployed for said uses;

2. Implementation of Accountability Protocols that hold Defendants responsible for Product actions and outputs and barred from further commercial deployment absent the Products' ability to follow a code of human-like ethical principles and guidelines and respect for human values and rights, and until Plaintiff and Class Members are fairly compensated for the stolen data on which the Products depend;

3. Implementation of effective cybersecurity safeguards of the Products as determined by the AI Council, including adequate protocols and practices to protect Users' PHI/PII collected through Users' inputting such information within the Products as well as through Defendants' massive web scraping, consistent with the industry standards, applicable regulations, and federal, state, and/or local laws;

4. Implementation of Appropriate Transparency Protocols requiring Defendants to clearly and precisely disclose the data they are collecting, including where and from whom, in clear and conspicuous policy documents that are explicit about how this information is to be stored, handled, protected, and used;

5. Requiring Defendants to allow Product users and everyday internet

users to opt out of all data collection and stop the illegal taking of internet data, delete (or compensate for) any ill-gotten data, or the algorithms which were built on the stolen data;

6. Requiring Defendants to add technological safety measures to the Products that will prevent the technology from surpassing human intelligence and harming others;

7. Requiring Defendants to implement, maintain, regularly review and revise as necessary, a threat management program designed to appropriately monitor Defendants' information networks for threats, both internal and external, and assess whether monitoring tools are appropriately configured, tested, and updated;

8. Establishment of a monetary fund (the "AI Monetary Fund" or "AIMF") to compensate class members for Defendants' past and ongoing misconduct to be funded by a percentage of gross revenues from the Products;

9. Appointment of a third-party administrator (the "AIMF Administrator") to administer the AIMF to members of the class as "data dividends" as fair and just compensation for the stolen data on which the Products depend;

10. Confirmation that Defendants have deleted, destroyed, and purged the PII/PHI of all relevant class members unless Defendants can provide reasonable justification for the retention and use of such information when weighed against the privacy interests of class members; and

11. Requiring all further and just corrective action, consistent with permissible law and pursuant to only those causes of action so permitted;

J. Award Plaintiff and the Class members their reasonable costs and expenses incurred in this Action, including attorneys' fees, costs, and expenses; and

K. Grant Plaintiff and the Class Members such further relief as the Court deems appropriate.

**JURY TRIAL DEMANDED**

Plaintiff demands a jury trial on all triable issues.

DATED: February 27, 2024

**GLANCY PRONGAY & MURRAY LLP**
*/s/Kevin F. Ruf*

KEVIN F. RUF (#136901)
1925 Century Park East, Suite 2100
Los Angeles, CA 90067
Telephone: (310) 201-9150
Facsimile: (310) 201-9160
Email: info@glancylaw.com

**GLANCY PRONGAY & MURRAY LLP**
Brian P. Murray
230 Park Avenue, Suite 358
New York, NY 10169
Tel: (212) 682-5340
Fax: (212) 884-0988
bmurray@glancylaw.com

**LAW OFFICE OF PAUL C. WHALEN**
Paul C. Whalen
768 Plandome Road
Manhasset, NY 11030
Tel: (516) 426-6870
pcwhalen@gmail.com

*Attorneys for Plaintiff*