MORGAN & MORGAN COMPLEX LITIGATION GROUP
Michael F. Ram (SBN 104805)
mram@forthepeople.com
711 Van Ness Ave, Suite 500
San Francisco, CA 94102
Tel.: (415) 358-6913

MORGAN & MORGAN COMPLEX LITIGATION GROUP
John A. Yanchunis (*pro hac vice*)
jyanchunis@forthepeople.com
Ryan J. McGee (*pro hac vice*)
rmcgee@forthepeople.com
201 N. Franklin St., 7th Floor
Tampa, FL 33602
Tel.: (813) 223-5505

CLARKSON LAW FIRM, P.C.
Ryan J. Clarkson (CA SBN 257074)
rclarkson@clarksonlawfirm.com
Yana Hart (CA SBN 306499)
yhart@clarksonlawfirm.com
Tiara Avaness (CA SBN 343928)
tavaness@clarksonlawfirm.com
Valter Malkhasyan (CA SBN 348491)
vmalkhasyan@clarksonlawfirm.com
22525 Pacific Coast Highway
Malibu, CA 90265
Tel: (213) 788-4050

CLARKSON LAW FIRM, P.C.
Tracey Cowan (CA SBN 250053)
tcowan@clarksonlawfirm.com
95 3rd St., 2nd Floor
San Francisco, CA 94103
Tel: (213) 788-4050

Counsel for Plaintiffs and the Proposed Classes

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

PLAINTIFFS MARILYN COUSART; NICHOLAS GUILAK; PAUL MARTIN; BREONNA ROBERTS; CAROLINA BARCOS; JAIR PAZ; ALESSANDRO DE LA TORRE; VLADISSLAV VASSILEV; SEAN ALEXANDER JOHNSON; JENISE MCNEAL; N.B, a minor; LORENA MARTINEZ; JOHN HAGAN, individually, and on behalf of all others similarly situated,

Plaintiffs,

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.: 23-cv-04557-VC

CLASS ACTION COMPLAINT

1. VIOLATION OF ELECTRONIC COMMUNICATIONS PRIVACY ACT, 18 U.S.C. §§ 2510, *et seq.*
2. VIOLATION OF THE COMPREHENSIVE COMPUTER DATA ACCESS AND FRAUD ACT ("CDAFA"), CAL. PENAL CODE § 502, *et seq*.
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. VIOLATION OF ILLINOIS'S BIOMETRIC INFORMATION PRIVACY ACT, 740 ILCS 14/1, *et seq.*

6. ILLINOIS CONSUMER FRAUD AND DECEPTIVE BUSINESS PRACTICES ACT 815 ILL. COMP STAT. §§ 505, *et seq.*

7. ILLINOIS CONSUMER FRAUD AND DECEPTIVE BUSINESS PRACTICES ACT 815 ILL. COMP. STAT. §§ 510/2, *et seq.*

8. NEGLIGENCE

9. INVASION OF PRIVACY

10. CONVERSION

11. UNJUST ENRICHMENT

DEMAND FOR JURY TRIAL

TABLE OF CONTENTS

Plaintiffs Marilyn Cousart; Nicholas Guilak; Paul Martin; Breonna Roberts; Carolina Barcos; Jair Paz; Alessandro De La Torre; Vladisslav Vassilev; Sean Alexander Johnson; Jenise McNeal; N.B., a minor; Lorena Martinez; and John Hagan (hereinafter "**Plaintiffs**"), individually and on behalf of all others similarly situated, bring 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**"). Plaintiffs' allegations are based upon personal knowledge as to themselves and their own acts, and upon information and belief as to all other matters based on the investigation conducted by and through Plaintiffs' 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."[1] 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 percent chance of crashing and killing everyone on board.[2] Humanity is now faced with the two Frostian roads Professor Hawking predicted we would have to choose between: One leads to sustainability, security, and prosperity; the other leads to civilizational collapse.

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

[2] 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.").

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,[3] 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."[4]

5. Defendants' Products, and the technology on which they are built, undoubtedly 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. Partnering with Microsoft, it doubled down on a strategy to secretly harvest massive amounts of personal data

[3] 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.

[4] 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.

from the internet, including private 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. Defendants 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 trained on stolen data, Defendants saw the immediate profit potential and rushed the Products to market without implementing proper privacy safeguards or other controls to ensure the Products 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.[5] Now again, we face imminent and unreasonable risks of the very fabric of our

[5] 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*

society 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.'"[6]

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[7] 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

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 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).

[6] *2021 Final Report*, NAT. SEC. COMM. ON A.I., https://cybercemetery.unt.edu/nscai/20211005220330/https://www.nscai.gov/ (last visited Dec. 22, 2023).

[7] 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.

close."[8] He is not alone.[9]

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.[10] 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***.[11]

---

[8] 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.

[9] 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.

[10] *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.

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

14. The head of the alignment team and safety at Open AI directly acknowledges these risks, postulating, "[b]efore 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, and we don't understand how it works. If we are not careful, we are setting ourselves up for a lot of correlated failures."[12]

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***?"[13]

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, such as IP addresses and location, including geolocation of the users; (10) social media information; (11) chat log data; (12) usage data; (13) analytics; (14) cookies;[14] (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,[15] user financial information through Stripe, musical tastes and preferences

[12] *Id*.; *see also* Jan Leike, @janleike, X (May 17, 2023, 10:56 AM), https://twitter.com/janleike/status/1636788627735736321.
[13] *Spotlight: AI Myths and Misconceptions—Transcript*, *supra* note 11.
[14] *Privacy Policy*, OPENAI, https://openai.com/policies/privacy-policy (last updated June 23, 2023).
[15] 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-

through Spotify,[16] user patterns and private conversation analysis through Slack and Microsoft Teams,[17] and even private health information obtained through the management of patient portals such as MyChart.[18] This information is captured in real time.

17. 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.[19]

18. Defendants must be enjoined from their ongoing violations of the privacy and property rights of millions and required to take immediate action to implement proper safeguards

---

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

[16] 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).

[17] 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 Dec. 22, 2023); 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).

[18] 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.

[19] 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.

and regulations for the Products, their users, and all of society, such as:

(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 safeguard or delete the misappropriated data, including the personal and sensitive information of millions on which they run, and follow a code of human-like ethical principles and guidelines and respect for human values and rights, and until Plaintiffs and Class Members are fairly compensated for the stolen property 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, and from continuing to divulge—on demand—the personal and private information of Product users and the millions of everyday internet users whose data Defendants misappropriated to build the Products.

PARTIES

Plaintiff Marilyn Cousart ("Plaintiff Cousart")

19. Plaintiff Cousart is and at all relevant times was a resident of the State of California.

20. Plaintiff Cousart started using ChatGPT-3.5 and ChatGPT-4.0 in 2023 from her personal computer for personal inquiries.

21. Plaintiff Cousart is a frequent user of various websites and social media platforms

which were scraped by Defendants, including Facebook, where she frequently shares content relating to personal life updates, her family, friends, trips, events, and food. She belongs to various Facebook groups such as marketplace groups for selling items, and groups relating to San Francisco history, relationships, gardening, and cooking. Plaintiff Cousart was caretaker to her father when he had cancer, and she frequently posted his private medical information and cancer experiences to purposely limited audiences on Facebook, including Facebook groups tailored to specific purposes and audiences, creating dedicated spaces where members can share insights, seek advice, and offer support with an expectation of privacy. Plaintiff Cousart reasonably expected that her posts and interactions within these and other restricted online groups would not be intercepted by any third-party. Had Plaintiff Cousart been aware that her posts and interactions were subject to the illegal data scraping practices described in this Complaint, by unauthorized third parties in violation of terms of service which reasonably assured her of the ongoing control and ownership of her data, including the right to delete such data, she would have refrained from participating in such discussions.

22. In addition to Facebook, Plaintiff Cousart also uses Instagram where she has posted content of herself, her family, friends, and her music. She has two Instagram accounts and uses them to post daily about her personal life and music. Plaintiff Cousart also has a Snapchat account that she uses for photo and video.

23. Plaintiff Cousart uses YouTube frequently and has posted her own videos to the platform, including videos featuring her face and voice. Plaintiff Cousart also has a Twitter and TikTok account for personal use and research purposes.

24. Plaintiff Cousart also uses Spotify to create unique playlists and interact with other people's playlists. She has an artist account and has posted a few of her songs to the platform.

25. Plaintiff Cousart also uses Dropbox, which contains pictures of her family and which she expected would only be accessed by a restricted audience. Plaintiff Cousart did not consent to having Defendants scrape her or her family's pictures/faces from Dropbox to train Defendants' AI Products and forever embed them into AI technology which may be used to create

digital clones.

26. Plaintiff Cousart reasonably expected that the information that she exchanged with these websites prior to 2021 would not be intercepted by any third-party looking to compile and use all her information and data for commercial purposes. Plaintiff Cousart did not consent to the use of her private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Cousart's personal data from across this wide swath of online applications and platforms to train the Products.

27. Plaintiff Cousart is concerned that Defendants have taken her personal information and statements, as reflected in her online contributions, and is also concerned about the misuse of her photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of herself and her family. Due to Defendants' illegal interference with her personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Cousart no longer has full control over that property, including her guaranteed legal right to delete it.

28. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything she shares online, Plaintiff Cousart's distress is exacerbated by the unacceptable dilemma she now faces: either surrender her and her family's personal information and privacy to Defendants without consent or compensation or forego the use of internet entirely.

Plaintiff Nicholas Guilak ("Plaintiff Guilak")

29. Plaintiff Guilak is and at all relevant times was a resident of the State of California.

30. Plaintiff Guilak first used ChatGPT-3.5 in or around March of 2023. Plaintiff Guilak uses ChatGPT-3.5 from his personal cell phone as well as both his work and personal computers.

31. Prior to 2021, Plaintiff Guilak engaged with a variety of websites and social media platforms which were scraped by Defendants, including posting acting videos and tutorials on Facebook and Instagram. On Facebook, he also frequently posts photos and videos of family

members, including his nieces and nephews, and comments on other users' content. Additionally, on several occasions, Plaintiff Guilak has posted information about his religious and political views.

32. Additionally, Plaintiff Guilak is also a frequent user of YouTube, where he maintains an active channel dedicated to acting, and provides tutorials on acting. Plaintiff has also posted videos and "demo reels" of his own auditions, which include his face and voice. Plaintiff Guilak did not consent to having Defendants scrape his voice or face from YouTube to train Defendants' AI Products and forever embed them into AI technology which may be used to create digital clones.

33. Plaintiff Guilak comments on Reddit; posting videos, pictures, and tweets on Twitter; posting videos and comments on TikTok; and posting and commenting on other users' accounts on Snapchat. Plaintiff Guilak uses his Spotify account to listen to music and create unique playlists.

34. In addition to personal use, Plaintiff Guilak also used a variety of these platforms to engage in professional self-promotion as an actor and to post teaching material for his students. This included sharing a great deal of personal content, such as photos and videos of auditions, performances, and training sessions. Moreover, Plaintiff Guilak has his own website, which hosts his headshots, clips, resume, demo reels, show reels, voice reels, and acting tips. Plaintiff Guilak regularly updates his online content including deleting content he no longer wishes to share with anyone.

35. Plaintiff Guilak is an active user of various Microsoft applications, including Microsoft Outlook, Word, Whiteboard, Notebook, PowerPoint, OneDrive, Teams, Microsoft Edge, Bing Chat, and Azure. These tools are crucial in Plaintiff Guilak's daily life, enabling him to manage email communications on Outlook, create and edit documents on Word, utilize digital whiteboarding for brainstorming and planning on Whiteboard, maintain notes and information on his professional career and self-promotion efforts on Notebook, develop presentations on PowerPoint, store and access data on OneDrive, collaborate with colleagues and friends on Teams,

and internet browse on Edge. Plaintiff Guilak also uses Bing Chat to interact with an AI-powered chat for quick information research, and Microsoft Azure to leverage cloud computing services for his professional needs. Given Plaintiff Guilak's extensive engagement with these platforms, a significant amount of his personal and sensitive information was exchanged across these Microsoft platforms.

36. Plaintiff Guilak is concerned that Defendants have taken his skills and expertise, as reflected in his online contributions, and incorporated it into Products that could someday result in professional obsolescence in actors and teachers like him.

37. Plaintiff Guilak reasonably expected that the information that he exchanged with these websites prior to 2021 would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff Guilak did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Guilak's personal data from across this wide swath of online applications and platforms to train the Products.

38. Plaintiff Guilak is concerned about the misuse of his photos, online contributions, and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of himself and his network of friends and family. Due to Defendants' illegal interference with his personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Guilak no longer has full control over that property, including his guaranteed legal right to delete it.

39. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff Guilak's distress is exacerbated by the unacceptable dilemma he now faces: either surrender his and his family's personal information and privacy to Defendants without consent or compensation or forego the use of internet entirely.

Plaintiff Paul Martin ("Plaintiff Martin")

40. Plaintiff Martin is and at all relevant times was a resident of the State of California.

41. Plaintiff Martin is a director of information technology and software engineer and

began using ChatGPT-3.5 on or about February/March 2023. He is a current user of ChatGPT-3.5 and ChatGPT-4.0. Plaintiff Martin accesses the Products from his personal computer, cellular device, and work computer.

42. Prior to 2021 and continuing to present day, Plaintiff Martin engages with a variety of websites and social media applications which were scraped by Defendants. Plaintiff Martin has had a Twitter account since approximately 2011; using it to post content, and re-post other users' tweets to save and compile information in line with his interests. For example, Plaintiff Martin has posted pictures of a concert he was attending with the location, song title of a song and even his friend's name.

43. For many years, Plaintiff Martin had a Spotify account which he frequently used to listen to music and create unique playlists. Approximately five (5) years ago, he transitioned to YouTube music and Google Play. Prior to 2021, Plaintiff Martin regularly viewed videos on YouTube, posted content such as a trailer video for a fictitious movie, and commented on other users' videos. Prior to 2021, he had a Facebook, Snapchat, and Instagram account. Plaintiff Martin published many posts on his Instagram account, which featured his face and voice and were accompanied by commentary. Plaintiff Martin did not consent to having Defendant scrape his voice or face to train Defendants' AI Products and forever embed them into AI technology that may be used to create digital clones.

44. Plaintiff Martin has posted photos of himself, his family, and friends on various websites and social media applications, including photos of his children and grandmother. He posted photos of himself and friends on online dating websites, such as OK Cupid and Tinder, approximately eight (8) years ago. He used these dating websites to meet potential romantic partners, and as a result disclosed significant amounts of personal information and exchange messages with prospective romantic partners. He has been using the United Healthcare Insurance Company web portal for over a decade to find providers and review post-appointment works.

45. Plaintiff Martin has also posted online about his political views, as well as frequently asked and answered technical questions using his professional knowledge on Stack

Overflow and GitHub for the last five (5) years in sporadic sprints to accumulate points on the website.

46. Plaintiff Martin is also an active user of the following Microsoft Services, including Visual Studio Subscription, Azure, OneDrive, Microsoft Office, Microsoft Edge, and has interacted with Bing Chat. These platforms are an integral part of Plaintiff Martin's daily activities, encompassing functions such as accessing a development tools and resources for software creation on Visual Studio Subscription, utilizing cloud computing and data management services on Azure, storing and accessing both personal and professional data in OneDrive, creating and managing documents, spreadsheets, and presentations for various task with Microsoft Office, internet browsing on Edge, and engaging with an AI-powered chat interface on Bing Chat.

47. Plaintiff Martin is concerned that Defendants have taken his skills and expertise, as reflected in his online contributions and incorporated them into Products that could someday result in professional obsolescence for software engineers like him.

48. Plaintiff Martin is also concerned that Defendants' practice of aggregating disparate pieces of personal information from multiple sources allows Defendants to form a comprehensive and exploitable profile of his identity. Specifically, Plaintiff Martin is concerned about his increased risk of identity theft and credit fraud, which poses a direct threat to his present financial decision making, security and privacy.

49. Plaintiff Martin reasonably expected that the information that he exchanged with these websites prior to 2021 would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff Martin did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Martin's personal data from across this wide swath of online applications and platforms to train the Products.

50. Plaintiff Martin is concerned about the misuse of his photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of himself and his network of friends and family. Due to Defendants' illegal interference

with his personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Martin no longer has full control over that property, including his guaranteed legal right to delete it.

51. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff Martin's distress is only exacerbated by the unacceptable dilemma he now faces: either surrender his personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff Breonna Roberts ("Plaintiff Roberts")

52. Plaintiff Roberts is and at all relevant times was a resident of the State of Illinois.

53. Plaintiff Roberts started using ChatGPT-3.5 in mid-2022 from her personal computer for general inquiries.

54. Plaintiff Roberts is a frequent user of various websites and social media platforms which were scraped by Defendants, including Tik Tok, where she frequently shares content relating to personal updates, posts about her spiritual business, discussions on healing from trauma, meditation practices, and personal experiences related to her life and relationships. She further uses TikTok to post videos of her daily life, which almost always feature her face and voice. Plaintiff Roberts did not consent to having Defendants scrape her voice or face to train Defendants' AI Products and forever embed them into AI technology that may be used to create digital clones.

55. In addition to professional and personal use on Tik Tok, Plaintiff Roberts also uses Instagram and Facebook, where she has posted pictures of herself, her friends, and her family. She actively uses the Facebook "groups" feature to post and re-post content about hair styling and cooking. At times, on her Facebook account and on these group channels, Plaintiff Roberts shares her personal experiences and spiritual health. Plaintiff Roberts posted and interacted with these groups believing they are tailored to specific purposes and restricted audiences. Plaintiff Roberts reasonably expected her posts and interactions within these groups would not be intercepted by any third-party. Had Plaintiff Roberts been aware that her posts and interactions were subject to data scraping practices by unauthorized third parties, she would have refrained from participating

in such discussions. Plaintiff Roberts reasonably expected that this information was shared strictly with limited audiences and did not expect Defendants to be using such information to develop their Products.

56. Plaintiff Roberts also uses Spotify to create unique playlists and interact with other people's playlists and YouTube to post pictures of personal vlogs of her daily experiences, videos of photoshoots, and clothing. Plaintiff Roberts also has a Yelp account which she has used multiple times to post comments and reviews on local businesses she has visited.

57. Plaintiff Roberts is a daily user of Microsoft Outlook, which she uses to manage emails, organize her calendars, schedule meetings and appointments, maintain task lists, and store contacts.

58. Plaintiff Roberts reasonably expected that the information that she exchanged with these websites prior to 2021 would not be intercepted by any third-party looking to compile and use all her information and data for commercial purposes. Plaintiff Roberts did not consent to the use of her private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Roberts's personal data from across this wide swath of online applications and platforms to train the Products.

59. Plaintiff Roberts is concerned that Defendants have taken her skills and expertise, as reflected in her online contributions, and incorporated it into Products that could someday result in professional obsolescence in giving spiritual advice. Plaintiff Roberts is also concerned about the misuse of her photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of herself and her network of friends and family. Due to Defendants' illegal interference with her personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Roberts no longer has full control over that property, including her guaranteed legal right to delete it.

60. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything she shares online, Plaintiff Roberts's distress is only exacerbated by the unacceptable dilemma she now faces: either surrender her and her family's

personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff Carolina Barcos ("Plaintiff Barcos")

61. Plaintiff Barcos is and at all relevant times was a resident of the State of California.

62. Plaintiff Barcos started using ChatGPT-3.5 in or around March of 2023. Plaintiff Barcos uses ChatGPT-3.5 from her personal cell phone as well as both her work and personal computers.

63. As an actor and a professor, Plaintiff Barcos maintains an active internet presence, commonly using platforms which were scraped by Defendants. For example, Plaintiff Barcos frequently uses Facebook and Instagram to engage in self-promotion and post teaching material, including sharing content, such as auditions, performances, and training sessions which feature her face and voice. Moreover, to spread awareness within these social networks, Plaintiff Barcos also posts media related to "psychological support," such as motivational quotes to cancer victims, and posts about reducing and preventing animal abuse. Plaintiff Barcos has also used Facebook to share many of her personal cooking recipes with friends and family.

64. Plaintiff Barcos is a member of a Facebook group tailored towards dog owners and dog lovers, in which she frequently shares photos and information about her dog. Plaintiff Barcos posted and interacted with this group reasonably believing it is tailored to a specific community of dog lovers. Had she been aware that her posts and interactions were subject to data scraping practices by unauthorized third parties, she would have refrained from posting in this group.

65. Plaintiff Barcos also uses Twitter to post text updates, photos, and videos; YouTube to share personal content and engage with other users in video comments; as well as TikTok, and Snapchat. Plaintiff Barcos has posted many photos of family members, including her nieces and nephews on these social media platforms. Plaintiff Barcos also uses Yelp to contribute her thoughts and commentary on local businesses.

66. Plaintiff Barcos is also an active user of the following Microsoft Services, including Microsoft Outlook, Microsoft Word, Microsoft Whiteboard, Microsoft Notebook, Microsoft PowerPoint, Microsoft One Drive, Microsoft Teams and Microsoft Edge. These platforms are an

integral part of Plaintiff Barcos' daily activities including managing communications via emails, crafting professional documents and reports, organizing and collaborating on projects with friends and colleagues, designing PowerPoint presentations for work, securely storing and accessing personal and professional data, coordinating team meetings and discussions, as well as browsing and researching information on the internet.

67. Plaintiff Barcos is concerned that Defendants have taken her skills and expertise, as reflected in her online contributions, and incorporated it into Products that could someday result in professional obsolescence for professors and educators like her.

68. Plaintiff Barcos reasonably expected that the information that she exchanged with these websites prior to 2021 and currently would not be intercepted by any third-party looking to compile and use all her information and data for commercial purposes. Plaintiff Barcos did not consent to the use of her private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Barcos's personal data from across this wide swath of online applications and platforms to train the Products.

69. Plaintiff Barcos is concerned about the misuse of her photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of herself and her network of friends and family. Due to Defendants' illegal interference with her personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Barcos no longer has full control over that property, including her guaranteed legal right to delete it.

70. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything she shares online, Plaintiff Barcos's distress is exacerbated by the unacceptable dilemma she now faces: either surrender her and her family's personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff Jair Paz ("Plaintiff Paz")

71. Plaintiff Paz is and at all relevant times was a resident of the State of California.

72. Plaintiff Paz started using ChatGPT-3.5 in or around December 2022 from his

personal account. Plaintiff Paz primarily uses ChatGPT for school, for example, to summarize study material. He also used and continues to use ChatGPT to create travel itineraries, inquire about sensitive health issues he has experienced, and investigate sensitive pet health issues. He accessed ChatGPT from his personal computer and cell phone, as well as his friend and sister's personal devices.

73. Plaintiff Paz has an active digital footprint, engaging with many websites and social media platforms which were scraped by Defendants. Plaintiff Paz has used YouTube to post videos sometime between 2016-2017. These videos include recordings of online video games such as MineCraft which contain his and other users' real-time voices as well as the faces of minors. Plaintiff Paz did not consent to having Defendants scraped his voice to train Defendants' AI Products and forever embed them into AI technology that may be used to create digital clones.

74. Further, Plaintiff Paz frequently uses Twitter, where he engages with other users' posted content, regarding topics such as his hobbies, interest, and political views. Plaintiff Paz uses Snapchat daily, sending photos of himself and using the application to communicate with friends; he made his account in or around 2016, when he was a minor.

75. Recently, when using Snapchat, Snapchat's "MyAI" automatically appeared in a group chat with Plaintiff Paz and his friends. Plaintiff Paz used it a few times, and then noticed it began generating responses that utilized past group chat content. Concerned that the "MyAI" chatbot had read and analyzed all of their conversations without their knowledge, Plaintiff Paz ended his use of "MyAI."

76. Plaintiff Paz also uses Instagram daily to message friends, share content with them via direct message, and frequently post images including his and his friends' faces. He also uses Reddit and has commented on other users' posts. He engages with content on TikTok, liking and sharing posts by other users either via the application itself or by posting links to TikTok via Discord or text message. Plaintiff Paz additionally uses Spotify to create playlists and interact with playlists created by other users.

77. Plaintiff Paz is also an active user of the following Microsoft Services, including

Microsoft Office 365, Microsoft Word, Microsoft Excel, Microsoft Outlook, Microsoft, PowerPoint, Microsoft Bing Chat, and Microsoft Edge. Plaintiff Paz uses Microsoft Word for drafting and editing documents, Microsoft Excel for managing and analyzing data through spreadsheets, Microsoft Outlook for email communication and calendar management, Microsoft PowerPoint for creating presentations for work and Microsoft Edge for internet browsing. Moreover, Plaintiff Paz frequently uses Bing Chat for real-time information retrieval and AI-assisted inquiries, which aids in quick information gathering.

78. Plaintiff Paz reasonably expected that the information that he exchanged with these websites prior to 2021 would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff Paz did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff J.P.'s personal data from across this wide swath of online applications and platforms to train the Products.

79. Plaintiff Paz is concerned about the misuse of his photos, online contributions and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of himself and his network of friends and family. Due to Defendants' illegal interference with his personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Paz no longer has full control over that property, including his guaranteed legal right to delete it.

80. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff Paz's distress is exacerbated by the unacceptable dilemma he now faces: either surrender his and his family's personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff Alessandro De La Torre ("Plaintiff De La Torre")

81. Plaintiff De La Torre is and at all relevant times was a resident of the State of California.

82. Plaintiff De La Torre is a product engineer and began using ChatGPT-3.5 on or

about November 2022. He is a current user of ChatGPT-3.5 and ChatGPT-4.0. Plaintiff De La Torre accesses the Products from his personal computer, cellular device, and work computer.

83. Prior to 2021 and continuing to present day, Plaintiff De La Torre engages with a variety of websites and social media applications which were scraped by Defendants. For example, Plaintiff De La Torre has accounts on Twitter, Reddit, TikTok, Snapchat, Yelp, LinkedIn, as well as Crunchbase, Webflow, and other technology-focused sites. Plaintiff De La Torre uses these platforms to post about a variety of topics, accompanied by commentary and visuals including his face, voice and location. Specifically, Plaintiff De La Torre has posted photos of himself, his cat, family members and friends on Instagram, some of which has included his location. Plaintiff De La Torre did not consent to having Defendants scrape his voice or face to train Defendants' AI Products and forever embed them into AI technology that may be used to create digital clones.

84. Plaintiff De La Torre has posted content on Twitter sharing his opinions and thoughts on current events, including the rapid development of artificial intelligence technology. Plaintiff De La Torre also uses TikTok to frequently post videos he has created encouraging his friends and family to take more risks in order to live a more fulfilling life.

85. For many years, Plaintiff De La Torre has had a Spotify account which he frequently uses to listen to music and create unique playlists. Plaintiff De La Torre regularly views videos on YouTube, posted content about application design and function, and commented on other users' videos.

86. Plaintiff De La Torre has also founded or co-founded at least four companies, the details of which are summarized on those respective websites.

87. Plaintiff De La Torre has also posted online about his political views, as well as frequently asked and answered technical questions using his professional knowledge on various websites such as LinkedIn. Plaintiff De La Torre uses LinkedIn for professional networking, using it to connect with colleagues and industry peers, seek and post job opportunities, engage with professional content, and participate in industry-specific discussions and groups.

88. Plaintiff De La Torre is concerned that Defendants have taken his skills and expertise, as reflected in his online contributions, and incorporated them into Products that could someday result in professional obsolescence for software engineers like him. Plaintiff De La Torre reasonably expected that the information that he exchanged with these websites would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff De La Torre did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff De La Torre's personal data from across this wide swath of online applications and platforms to train the Products.

89. Plaintiff De La Torre is deeply concerned about the misuse of his photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of himself and his network of friends and family. Due to Defendants' illegal interference with his personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff De La Torre no longer has full control over that property, including his guaranteed legal right to delete it.

90. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff De La Torre's distress is exacerbated by the unacceptable dilemma he now faces: either surrender his personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff Vladisslav Vassilev ("Plaintiff Vassilev")

91. Plaintiff Vassilev is and at all relevant times was a resident of the State of California.

92. Plaintiff Vassilev started using ChatGPT-3.5 and ChatGPT-4.0 in late 2022 from his personal computer and cellphone for general inquiries.

93. Plaintiff Vassilev is a frequent user of various websites and social media platforms, including Reddit, where he posts questions and content related to his knowledge of video games.

94. Plaintiff Vassilev uses Instagram and shares content relating to personal updates, family, travel, vacations, and events he attends. He has shared photos of his family, fiancé, and

daughter, featuring his face and voice on many of the posts. Plaintiff Vassilev did not consent to having Defendants scrape his voice or face to train Defendants' AI Products and forever embed them into AI technology that may be used to create digital clones.

95. Plaintiff Vassilev also uses Reddit to post questions and inquiries relating to video games and Yelp to post reviews on local restaurants.

96. Plaintiff Vassilev also uses Spotify to listen to music, create unique playlists and interact with other people's playlists. He follows his favorite musical artists and interacts with their playlists.

97. Plaintiff Vassilev reasonably expected that the information that he exchanged with these websites prior to 2021 and currently would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff Vassilev did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Vassilev's personal data from across this wide swath of online applications and platforms to train the Products.

98. Plaintiff Vassilev is concerned about the misuse of his photos, online contributions, and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of himself and his network of friends and family. Due to Defendants' illegal interference with his personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Vassilev no longer has full control over that property, including his guaranteed legal right to delete it.

99. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff Vassilev's distress is exacerbated by the unacceptable dilemma he now faces: either surrender his and his family's personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff Sean Alexander Johnson ("Plaintiff Johnson")

100. Plaintiff Johnson is and at all relevant times was a resident of the State of California.

101. Plaintiff Johnson first accessed used ChatGPT-3.5 in or around December of 2022.

Plaintiff Johnson accesses ChatGPT-3.5 from his personal computer and personal cell phone. He has used ChatGPT-3.5 many times for writing emails, creating stories, and performing educational research.

102. Prior to 2021, Plaintiff Johnson engaged with a variety of websites and social media applications which were scraped by Defendants, including Spotify, where he has created a channel and posted music across many different genres.

103. Plaintiff Johnson is also an active user of Microsoft Word, using it frequently to draft and edit documents for both personal and professional work purposes.

104. Plaintiff Johnson reasonably expected that the information that he exchanged with these websites would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff Johnson did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Johnson's personal data from across this wide swath of online applications and platforms to train the Products. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff Johnson's distress is exacerbated by the unacceptable dilemma he now faces: either surrender his personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff Jenise McNeal ("Plaintiff McNeal")

105. Plaintiff McNeal is and at all relevant times was a resident of the State of California.

106. Plaintiff McNeal started using ChatGPT-3.5 in or around August of 2023. Plaintiff McNeal uses ChatGPT-3.5 from her personal computer, mainly to research and obtain information on a wide variety of topics, including cooking, parenting and child development.

107. Plaintiff McNeal uses multiple platforms which were scraped by Defendants. For example, Plaintiff McNeal actively uses Facebook and Instagram to post family updates, consistently sharing content like pictures and videos featuring her cousins, grandmother, and her eleven-year-old daughter. Often these posts prominently display both her face and voice. Plaintiff

McNeal is a member of multiple Facebook groups, including a group called "I love San Bernandino"—where she has posted details about her mother's house fire and "Fashion Show Group"—where she has shared her personal outfits to engage in open discussions and commentary on fashion. As a member of these Facebook groups, Plaintiff McNeal reasonably expected that the information she shared would be strictly confined to the specific audiences of these groups, preserving the privacy of her posts.

108. Plaintiff McNeal also uses TikTok to upload videos and reels which showcase herself and her minor daughter cooking and creating funny dance videos. Moreover, Plaintiff McNeal has shared videos expressing her personal opinions and perspective, including insights into her personal life and reflections of her previous relationship.

109. Plaintiff McNeal uses various streaming services, namely Pandora, Soundcloud, and Apple Music. Plaintiff McNeal follows and interacts with bands and creates playlists.

110. Plaintiff McNeal reasonably expected that the information that she exchanged with these websites would not be intercepted by any third-party looking to compile and use all her information and data for commercial purposes. Plaintiff McNeal did not consent to the use of her private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff McNeal's personal data from across this wide swath of online applications and platforms to train the Products.

111. Plaintiff McNeal is concerned about the misuse of her and her minor's photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of herself and her daughter. McNeal is extremely concerned for her daughter's safety and privacy. Due to Defendants' illegal interference with her personal and family information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff McNeal no longer has full control over her property, including her guaranteed legal right to delete it.

112. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff McNeal's distress is exacerbated by

the unacceptable dilemma she now faces: either surrender her and her daughter's personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff N. B.

113. Plaintiff N.B. is and at all relevant times was a resident of the State of California.

114. Plaintiff N.B is an eleven (11) year old minor, who is currently in the sixth grade.

115. Plaintiff N.B started using ChatGPT-3.5 in or about August of 2023. Neither Plaintiff N.B.'s parents or guardian was required to review any terms and policies, including Terms and Policies or the Privacy Policy of OpenAI. N.B's parents and/or guardians did not review any such documents prior to N.B.'s creation and/or use of the OpenAI account at issue.

116. Plaintiff N.B is an active user of Instagram, regularly engaging with the platform to post personal content, including pictures and videos featuring her aunt, cousins, and herself. These posts prominently display her face and voice.

117. As a sixth grader, Plaintiff N.B also uses Microsoft Office services for many educational purposes, such as creating and editing documents in Word for writing assignments, utilizing PowerPoint for developing presentations for school projects, and accessing OneNote for note-taking and organizing study materials.

118. Plaintiff N.B and her guardian reasonably expected that the information that the minor exchanged with the Products and online would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff N.B. and her guardian did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff N.B.'s personal data to train the Products.

Plaintiff Lorena Martinez ("Plaintiff Martinez")

119. Plaintiff Martinez is and at all relevant times was a resident of the State of Florida.

120. Plaintiff Martinez first started using ChatGPT-3.5 on her personal computer. Plaintiff Martinez now also uses ChatGPT-4.0 on her personal cell phone.

121. Prior to 2021, Plaintiff Martinez actively used and to this day continues to use a wide variety of websites and social media platforms which were scraped by Defendants. Plaintiff

Martinez posts many photos on Facebook of her friends, her family, and her minor son. Plaintiff Martinez has posted hundreds of photos of her son on Facebook and uses Instagram to do the same, with slightly less frequency.

122. Plaintiff Martinez belongs to various Facebook groups relating to cleaning, household activities, single mother support, children with autism support, and private health communities where she has shared her thoughts and comments. Plaintiff Martinez has used these groups to interact with similarly situated individuals. Plaintiff Martinez is facing medical procedures, so she has posted personal information in various targeted and restricted Facebook support groups for women in similar situations. Plaintiff Martinez has posted and interacted with these groups believing they are tailored to specific purposes and audiences. Plaintiff Martinez reasonably expected her posts and interactions within these groups to be would not be intercepted by any third-party. Had Plaintiff Martinez been aware that her posts and interactions were subject to data scraping practices by unauthorized third parties, she would have refrained from participating in such discussions.

123. Plaintiff Martinez has a TikTok account, where she creates and shares videos, many of which feature her face and voice. Plaintiff Martinez also uses Spotify to create unique playlists and interact with other people's playlists. She further uses Reddit and Twitter to read articles and comment on other users' posts about topics she is interested in, including parenting advice, celebrity news, politics, and single motherhood support. Plaintiff Martinez also uses Yelp to post business reviews.

124. Plaintiff Martinez is also an active user of the following Microsoft Services, including Windows Authentication, Microsoft Outlook, Microsoft Office Suite, and Microsoft Teams. Plaintiff Martinez uses Microsoft Outlook and Microsoft Teams to communicate with friends and colleagues via email communications, manage her calendar, host and attend video conferences and organize her daily tasks.

125. Plaintiff Martinez reasonably expected that the information that she exchanged with these websites prior to 2021 would not be intercepted by any third-party looking to compile and

use all her information and data for commercial purposes. Plaintiff Martinez did not consent to the use of her private information by third parties in this manner. In fact, Plaintiff Martinez maintains her social media accounts with strict privacy settings, clearly expressing her intention to limit the sharing of her information to a small, private circle of people, thereby not consenting to any third-party use of her personal data. Notwithstanding, Defendants stole Plaintiff Martinez's personal data from across this wide swath of online applications and platforms to train the Products.

126. Plaintiff Martinez is concerned about the misuse of her photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of herself and her son. Due to Defendants' illegal interference with her personal and family information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Martinez no longer has full control over her property, including her guaranteed legal right to delete it.

127. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything she shares online, Plaintiff Martinez's distress is exacerbated by the unacceptable dilemma she now faces: either surrender her and her son's personal information and privacy to Defendants or forego the use of internet entirely.

Plaintiff John Hagan ("Plaintiff Hagan")

128. Plaintiff Hagan is and at all relevant times was a resident of the State of New York.

129. Plaintiff Hagan is a consumer and began using ChatGPT-3.5 on or about December 2022. He is a current user of ChatGPT-3.5 and ChatGPT-4.0. Plaintiff Hagan accesses the Products from his personal computer and cellular device.

130. Prior to 2021 and continuing to present day, Plaintiff Hagan engages with a variety of websites and social media which were scraped by Defendants. For example, Plaintiff Hagan frequently uses his Instagram and Twitter accounts to post photos of himself, family members and friends, many of which feature his face, voice, and location. Plaintiff Hagan also uses Twitter to share political posts as part of his internship, aiming to promote awareness of international affairs and poverty. Moreover, he uses Facebook to interact with his friends and family, as well as engage

in commerce on Facebook Marketplace. Plaintiff Hagan is part of certain Facebook groups and Reddit communities focused around posting advice and teaching tips on LSAT preparation. Plaintiff Hagan has posted and interacted with these groups reasonably believing they are tailored to specific purposes and audiences. Plaintiff Hagan would not have posted and interacted with these communities had he known that his contributions were subject to data scraping practices by unauthorized third parties.

131. For many years, Plaintiff Hagan has had a Spotify account which he frequently uses to listen to podcasts. Plaintiff Hagan regularly views videos on YouTube and has posted videos to YouTube. Plaintiff Hagan has a variety of other social media accounts including TikTok, Snapchat, Discord, and Yelp. Plaintiff Hagan published many posts on these internet accounts, accompanied by his personal thoughts and commentary.

132. Plaintiff Hagan has had other social media applications in the past but deleted them in December 2022, expecting that the information posted within those platforms would remain in his exclusive control and not be used without his authorization.

133. Plaintiff Hagan reasonably expected that the information that he exchanged with these websites would not be intercepted by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff Hagan did not consent to the use of his private information by third parties in this manner. Notwithstanding, Defendants stole Plaintiff Hagan's personal data from across this wide swath of online applications and platforms to train the Products.

134. Plaintiff Hagan did not consent to have his voice or face scraped into a database that can be used to create digital clones. Plaintiff Hagan is concerned about the misuse of his photos and private information, including having significant anxiety, distress, vulnerability and fear for the privacy and safety of himself and his network of friends and family. Due to Defendants' illegal interference with his personal information, and specifically embedding it permanently into AI Products and the models on which they run, Plaintiff Hagan no longer has full control over his property, including his guaranteed legal right to delete it.

135. Because Defendants offer no effective opt out from the ongoing misappropriation and commercialization of anything he shares online, Plaintiff Hagan's distress is exacerbated by the unacceptable dilemma he now faces: either surrender his personal information and privacy to Defendants or forego the use of internet entirely.

Defendants

136. **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").[20] OpenAI was founded in 2015 and is headquartered in San Francisco, CA. OpenAI has released the AI-based products DALL-E, GPT-4, OpenAI Five, ChatGPT, and OpenAI Codex for commercial (to integrate within one's business) and personal use.

137. 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."[21] 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."[22] Fast forward to April of 2023: OpenAI closed a more than $300 million share sale at a valuation between $27 billion and $29 billion.[23] 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.[24]

138. Defendant OpenAI GP, L.L.C. ("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, L.P. and is responsible for managing and operating the day-to-day business

[20] *OpenAI LP*, OPENAI, https://openai.com/blog/openai-lp (last visited Dec. 22, 2023).
[21] Greg Brockman & Ilya Sutskever, *Introducing OpenAI*, OPENAI (Dec. 11, 2015), https://openai.com/blog/introducing-openai.
[22] *Id.*
[23] *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/.
[24] 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/.

and affairs of OpenAI, L.P. Its primary focus is research and technology. OpenAI GP was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI GP is liable for the debts, liabilities, and obligations of OpenAI, L.P., including litigation and judgments.

139. Defendant OpenAI Startup Fund I, L.P. ("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, L.P., 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, L.P. OpenAI Startup Fund I exercised control over OpenAI, L.P. and was aware of the unlawful conduct alleged herein throughout the Class Period.

140. Defendant OpenAI Startup Fund GP I, L.L.C. ("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.

141. 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.

142. **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.[25] 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.[26] Then, in January of 2023, Microsoft confirmed its extended partnership with OpenAI by investing $10 billion into ChatGPT.[27] Prior to this $10 billion dollar investment, Microsoft had invested $3 billion into OpenAI in previous years.[28]

143. 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.[29] Furthermore, Microsoft Azure also acts as the exclusive cloud service of OpenAI.[30] Microsoft even sits on OpenAI's Board.

144. As OpenAI's largest investor and largest service provider—specifically in connection with the development of ChatGPT—Microsoft exerts considerable control over

[25] 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.

[26] 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.

[27] *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.

[28] 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.

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

[30] *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.

OpenAI. Analysts estimate OpenAI will add between $30 billion and $40 billion to Microsoft's top line.

145. **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

146. 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.

147. 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 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.

148. 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.

149. 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.

150. 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

151. 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. . . ."[31] 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."[32] 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.

152. OpenAI also originally pledged to "freely collaborate" with other responsible

[31] *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/.
[32] *Id.*

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."[33]

153. 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.

154. 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 power required.[34] 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 (on stolen data) that ultimately resulted in ChatGPT and the image generator DALL-E.[35]

155. 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.

---

[33] Greg Brockman & Ilya Sutskever, *Introducing OpenAI*, OPENAI (Dec. 11, 2015), https://openai.com/blog/introducing-openai.
[34] 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.
[35] *Id.*

156. 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.

157. 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."[36] 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.[37] Defendants, like 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.

---

[36] *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/.

[37] "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, *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/.

158. 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.

159. 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.[38] They worry that OpenAI is prioritizing short-term financial gains over long-term safety, legal, and ethical considerations, as exemplified by the surreptitious mass theft of personal information and sudden deployment of the Products for widespread commercial use, despite all the known dangers.[39] 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."[40]

160. Elon Musk, an original non-profit funder and founder, was blunter as to the 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."[41]

161. 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?"[42] This same question must

[38] *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.
[39] *Id.*
[40] 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.
[41] 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.
[42] Elon Musk, @elonmusk, X (Mar. 15, 2023), https://twitter.com/elonmusk/status/1636047019893481474.

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***.

162. Indeed, it appears this ethical dilemma has infected OpenAI's own internal leadership. On November 17, 2023, CEO Sam Altman was suddenly fired by OpenAI's board, who claimed that Altman was "not consistently candid in his communications with the board."[43] That same day, co-founder Greg Brockman announced that he was quitting, "based on today's news."[44] Tech Journalist Kara Swisher reported that, according to her sources, these major departures were the result of a "'misalignment' of the profit versus nonprofit adherents at the company."[45]

163. Complete chaos ensued at the company in the days to follow. On Saturday, a host of OpenAI's for-profit investors, including Microsoft, urged the company to rehire Altman.[46] OpenAI succumbed to the pressure and agreed, and Altman was back at OpenAI offices the following day.[47] But everything once again fell apart the evening of November 19th, when OpenAI announced that it was bringing on Twitch executive Emmett Shear as its new interim CEO.[48] On Monday the 20th, Microsoft announced that Brockman and Altman would be joining Microsoft to "lead a new advanced AI research team."[49]

164. Brockman was among the many OpenAI employees who were upset by Altman's

[43] Rachel Mitz, *Sam Altman Pushed Out at OpenAI After Board Loses Confidence*, L. A. TIMES (Nov. 17, 2023), https://www.latimes.com/business/story/2023-11-17/sam-altman-pushed-out-at-openai-after-board-loses-confidence.
[44] Max Zahn, *4 Days from Fired to Re-hired: A Timeline of Sam Altman's Ouster from OpenAI*, ABC NEWS (Nov. 22, 2023), https://abcnews.go.com/Business/sam-altman-reaches-deal-return-ceo-openai/story?id=105091534.
[45] Matt Binder, *Here's a Timeline of the OpenAI Saga with CEO Sam Altman*, MASHABLE (Nov. 22, 2023), https://mashable.com/article/openai-sam-altman-saga-timeline; Kara Swisher (@karaswisher), X (Nov. 17, 2023), https://twitter.com/karaswisher/status/1725678074333635028?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1725678074333635028%7Ctwgr%5Eb63df8387dbc086dc8404a513860926f7e67b2d9%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fmashable.com%2Farticle%2Fopenai-sam-altman-saga-timeline.
[46] Zahn, *supra* note 44.
[47] *Id.*
[48] *Id.*
[49] *Id.*

departure.[50] In light of the negative response to Altman's firing, OpenAI once again decided to bring Altman back as CEO on November 21st.[51] Twelve days after his initial ouster, OpenAI officially announced that it had decided to hire Altman back—for good this time.[52]

165. This implosion of OpenAI leadership demonstrates that within the company there is an awareness of the grave ethical concerns the technology poses, and the threat posed by the Company's focus on commercialization and putting profits over people, their property and privacy rights, and their safety. What is clear is that the public cannot trust this company navigating a clearly intense internal strain to make good on its vague promises of upstanding intentions for AI. Soon after the upheaval, Microsoft joined OpenAI's Board.

166. 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 partnership with Microsoft and machinations for profit, "the most powerful tool mankind has ever created, is now in the hands of a ruthless corporate monopoly."[53]

B. Microsoft's Was Directly Involved in Developing and Training the GPT Products and Has Profited off its Partnership with OpenAI

167. For the last several years, Microsoft has been deeply involved with the development, training, and commercialization of OpenAI's GPT Products. Microsoft has invested at least $13 billion in OpenAI, and reportedly owns a 49 percent stake in the company's for-profit

[50] Binder, *supra* note 45. "Nearly all 800 employees at OpenAI signed a letter calling for the resignation of the company's board and the return of Altman as CEO." *Id.* The employees said that if their demands were not met, they would all quit. *Id.* Even one of the board members responsible for ousting Altman, Ilya Sutskever, signed this letter and admitted that he "deeply regret[ted] [his] participation in the board's actions." *Id.*
[51] *Id.*
[52] David Godman, *OpenAI Officially Announces Sam Altman has Returned as CEO and Microsoft Gains a Non-Voting Board Seat*, CNN BUSINESS (Nov. 29, 2023), https://www.cnn.com/2023/11/29/tech/openai-sam-altman-board-microsoft/index.html.
[53] 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.

operations. Microsoft's CEO has referred to the company's relationship with OpenAI as a "great commercial partnership."

168. To develop the GPT programs, OpenAI needed a specialized supercomputing system. Microsoft's Azure provided the cloud computing systems necessary to train the GPT products, and Azure continues to power OpenAI operations to this day. Without the help of Azure, OpenAI would not have been able to develop and profit off its AI products. Through its creation and maintenance of the supercomputing system which was fed the stolen information at issue, Microsoft participated actively in the wrongdoing from the start, and continues to play an active role.

169. In addition to its role in facilitating the training process, Microsoft also played a key role in commercializing OpenAI's GPT products, and in doing so, has profited off of the technology and OpenAI's web-scraping practices.

170. Microsoft distributed and sold GPT-based products, like Azure products that incorporate GPT-3 and GPT-4, which it has also in turn profited off. Since incorporating GPT-3 into Bing search engine, Bing surpassed more than 100 million daily active users. Further, Microsoft has been integrating ChatGPT into Azure and Office 365 products and charging additional fees for the generative AI offerings. Analysts predict that Microsoft's integration of GPT systems into Microsoft Products could generate over $10 billion in annualized revenue by 2026.[54] In fact, one existing version of the integration, "GitHub Copilot," already generates more than $100 million in annual recurring revenue.[55] None of this revenue would have been possible

[54] Jordan Novet, *Microsoft Starts Selling AI Toll for Office, Which Could Generate $10 Billion a Year by 2026*, CNBC (Nov. 1, 2023), https://www.cnbc.com/2023/11/01/microsoft-365-copilot-becomes-generally-available.html.

[55] Aaron Holmes, *Microsoft's GitHub AI Coding Assistant Exceeds $100 Million in Recurring Revenue*, THE INFORMATION, https://www.theinformation.com/briefings/microsoft-github-copilot-revenue-100-million-ARR-ai (last visited Dec. 5, 2023).

without the stolen information at issue, and Plaintiffs and the Classes deserve compensation for this unjust enrichment.

C. OpenAI's Products

171. 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 stolen ("training") data.

172. ChatGPT was released as a "research preview" on November 30, 2022.[56] 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 users** in only two months, making it the fastest-growing app in history.[57] For comparison, TikTok took nine months to reach the same benchmark.[58] ChatGPT has continued to evolve exponentially, **with 1.8 billion visits in April and May of 2023**.[59]

173. ChatGPT was built on a family of large language models ("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,[60] 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."[61] 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.

---

[56] *Introducing ChatGPT,* OPENAI (NOV. 30, 2022), https://openai.com/blog/chatgpt.
[57] 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/.
[58] *Id.*
[59] Nerdynav, *97+ ChatGPT Statistics & User Numbers in June 2023 (New Data)*, NERDY NAV (June 2, 2023), https://nerdynav.com/chatgpt-statistics/.
[60] 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.
[61] Mark Wilson, *ChatGPT Explained: Everything You Need to Know About the AI Chatbot*, TECHRADAR (Mar. 15, 2023), https://www.techradar.com/news/chatgpt-explained.

174. 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. Microsoft and OpenAI also developed a program VALL-E, which has not been released for use to the public yet.

175. 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."[62] DALL-E uses a version of GPT-3, modified to generate images.[63]

176. OpenAI Five is a computer program developed by OpenAI that plays the five-on-five video game Dota 2.[64]

177. OpenAI Codex is another artificial intelligence model developed by Open AI, which is programmed to generate computer code for use in programming applications.[65]

178. 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.[66] VALL-E was illegally trained on audio voices from thousands of speakers without consent, notice, or compensation.[67]

D. ChatGPT's Development Depends on Secret Web-Scraping

179. 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.

---

[62] 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/.
[63] *Id.*
[64] 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/.
[65] Thomas Smith, *Why OpenAIs Codex Won't Replace Coders*, IEEE SPECTRUM (Sept. 28, 2021), https://spectrum.ieee.org/openai-wont-replace-coders.
[66] *VALL-E Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers*, GITHUB PAGES, https://lifeiteng.github.io/valle/index.html (last visited Dec. 22, 2023).
[67] *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.

180. 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.

181. 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."[68] 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 IV.A).

182. "Scraping involves the use of 'bots,' or robot applications deployed for automated tasks, which scan and copy the information on webpages then *store* and *index* the information."[69] 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."[70]

183. 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."[71] Thus, swept up into the Products is "a lot of data about you and me."[72] 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.[73] Many of these images were

[68] 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.
[69] 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/.
[70] 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.
[71] *Id.*
[72] *Id.*
[73] Hern & Milmo, *supra* note 70.

of "children and came from photo sites and personal blogs."[74]

184. 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 and all the content written by human beings."[75]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 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.[76] 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.[77] Defendants have scraped thousands of websites to collect this personal information. Plaintiffs have compiled

[74] Drew Harwell, *AI-generated child sex images spawn new nightmare for the web*, THE WASH. POST (June 19, 2023), https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images/.

[75] 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.

[76] *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.").

[77] 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.

a selection of around 1,000 websites that Defendants have scraped to illustrate the breadth and character of Defendants' scraping practices. *See* **Exhibit A** (Misappropriated Content – Representative List of Websites).

185. As reflected in **Exhibit A**, the breadth and scope of Defendants' data collection without permission, impacting essentially every internet user ever, raises serious legal, moral, and ethical issues.[78] 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."*[79] 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.[80]

186. Despite "*Open"* AI's "absolute secrecy" surrounding its data collections and practices,[81] we know at the highest levels that the Company used (at least) six (6) 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; (5) BookCorpus; and (6) Wikipedia.[82]

187. 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 Plaintiffs and the Classes' posted personal information, video, and audio clips of themselves and more, e.g., YouTube, Facebook, TikTok, Snapchat, and Instagram. Given Defendants' scraping

[78] 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.").
[79] Gal, *supra* note 68.
[80] Hern & Milmo, *supra* note 70.
[81] *Id.* ("Copyright lawsuits and regulator actions against OpenAI are hampered by the company's absolute secrecy about its training data.").
[82] 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.

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.

188. 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."[83] 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."[84]

189. Defendants' theft related to their WebTex2 corpus is ongoing and continuous. As one article explains that 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."[85] Neither Reddit itself nor Reddit users, much less all the owners of the webpages and personal data linked to and from Reddit, consent to this taking of data.

190. BookCorpus is a controversial dataset assembled in 2015 by copying the books from a website called Smashwords.com that hosts unpublished novels that are available to readers at no cost.[86] Those novels were copied into the BookCorpus dataset without consent, credit, or compensation to the authors. OpenAI misappropriated BookCorpus, knowing it was already once-pirated information, to train GPT-1, a collection of "over 7,000 unique unpublished books from a variety of genres…"[87] OpenAI admitted the value of this stolen data, stating, "[c]rucially, it contains long stretches of contiguous text, which allows the generative model to learn to condition

[83] 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 Nov. 15, 2023).
[84] *Id.*
[85] 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.
[86] Jack Bandy, *Dirty Secrets of BookCorpus, a Key Dataset in Machine Learning*, Medium (May 12, 2021), https://towardsdatascience.com/dirty-secrets-of-bookcorpus-a-key-dataset-in-machine-learning-6ee2927e8650.
[87] Radford et. Al., *Improving Language Understanding by Generative Pre-Training,* OPENAI, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (last accessed Dec. 5, 2023).

on long range information."[88]

191. OpenAI has further admitted that Books1 and Books2 datasets used to train ChatGPT are "internet-based books corpora."[89] Both Books1 and Books2 are much larger than BookCorpus.[90] Based on OpenAI's published information regarding GPT-3, Books1 is apparently about 9 times larger and Books2 is about 42 times larger.[91] Since BookCorpus contained about 7,000 titles, this suggests Books1 would contain about 63,000 titles and Books2 would contain about 294,000 titles. The only "internet-based books corpora" that have ever offered that much material are notorious "*shadow library*" websites like Library Genesis (aka LibGen), Z-Library (aka B-ok), Sci-Hub, and Bibliotik. The books aggregated by these websites have also been available in bulk via torrent systems.

192. Such illegal shadow libraries present great value to reckless AI companies seeking to train their LLMs. For example, an AI training dataset published in December 2020 by EleutherAI called "Books3" includes a recreation of the Bibliotik collection and contains nearly 200,000 books.[92] On information and belief, the OpenAI Books2 dataset includes books copied from these known "shadow libraries," as those are sources of trainable books most similar in nature and size to OpenAI's description of Books2. **Accordingly, OpenAI has used scraped, *pirated content* to train ChatGPT.**

193. Tellingly, **OpenAI has refused to even disclose the precise details of content used to train GPT-4,** broadly and evasively referring to the misappropriated data as "internet data."[93]

194. Another primary data set on which the Products depend, that the public currently

[88] *Id.*
[89] Tom Brown et al. *Language Models are Few-Shot Learners,* OPENAI (July 22, 2020), https://arxiv.org/pdf/2005.14165.pdf. Pg.
[90] *Id.*
[91] *Id.* (Figure 2.2 depicting 12 billion tokens in Books 1 and 55 billion tokens in Books2); *see also* Brandy, *supra* note 86 (BookCorpus amounting to 984 million words (i.e., approximately 1.3 billion tokens).
[92] Alex Perry, *A Giant Online Book Collection Meta Used to Train its AI is Gone Over Copyright Issues,* Mashable (Aug. 18, 2023), https://mashable.com/article/books3-ai-training-dmca-takedown.
[93] *GPT-4 Technical Report*, OPENAI (March 27, 2023), https://cdn.openai.com/papers/gpt-4.pdf.

knows about, is the "Common Crawl," a massive collection of web pages and websites also derived from large-scale web scraping and misappropriated by Defendants. It contains petabytes of data collected over twelve (12) years, including raw webpage data, metadata extracts, and text extracts from all types of websites.[94] In total, the Common Crawl dataset constitutes nearly a trillion words.

195. 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.[95] 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.[96] 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.

196. Over the course of eight (8) years, the Common Crawl dataset, used to train OpenAI's AI Products has scraped over 25 billion websites,[97] including countless high-traffic sites with privacy policies representing data security, terms of service promising data ownership and/or required passwords protection features.

197. **Defendants have scraped private websites with password protection and restricted access**. From just a sampling of the thousands+ websites Defendants scraped from 2018 to 2022 alone, hundreds are password protected. For example, facebook.com, Instagram.com, tiktok.com, whatsapp.com, spotify.com, reddit.com, outlook.com, twitter.com, dropbox.com,

[94] *Want to Use Our Data*, COMMON CRAWL, https://commoncrawl.org/the-data/ (last visited June 27, 2023).
[95] 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.
[96] 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.
[97] Ryan Elkins, *Search the html Across 25 Billion Websites for Passive Reconnaissance Using Common Crawl*, MEDIUM (Jul. 3, 2020), https://medium.com/@brevityinmotion/search-the-html-across-25-billion-websites-for-passive-reconnaissance-using-common-crawl-7fe109250b83.

stackoverflow.com, office.com, snapchat.com, linkedin.com, crunchbase.com, webflow.com, soundcloud.com, discord.gg, wordpress.com, pinterest.com, blogspot.com, yelp.com, and vimeo.com.

198. Plaintiff Cousart never expected that the content she shared on Facebook with her close network and specific audiences regarding caring for her father through his cancer experience would be scraped to train AI. Plaintiff Cousart also never expected that private photos of her family stored in her Dropbox account, or her photos posted to Instagram, to be scraped to train AI. Plaintiff Cousart also remains anxious and fearful that her and her family's faces can be misused to create digital clones.

199. Plaintiff Guilak never expected that the content he posted to Facebook, Snapchat, and Instagram, from photos of his family, nieces and nephews, to his religious and political views, would be scraped to train AI. Plaintiff Guilak also never anticipated that his comments on Reddit, his tweets posted to Twitter, videos and comments posted to TikTok, or his unique Spotify playlists would be scraped to train AI.

200. Plaintiff Martin never anticipated that his posts on Twitter, photos posted to Instagram, or his unique Spotify playlists would be scraped to train AI. Plaintiff Martin also never expected that questions he answered on Stack Overflow, utilizing his professional knowledge, would be scraped to train AI.

201. Plaintiff Roberts never anticipated that her photos posted to Instagram or Facebook, her content posted to specific Facebook groups, or her personal content for her followers concerning spiritual practices for healing trauma, would be scraped to train AI. Plaintiff Roberts also never expected that her Yelp reviews and comments would be scraped to train AI.

202. Plaintiff Barcos never anticipated that her content posted to Instagram, Twitter, TikTok, Snapchat, or Facebook, including her content posted to specific Facebook groups for psychological support to cancer victims, would be scraped to train AI. Plaintiff Barcos also never expected that her Yelp comments would be scraped to train AI.

203. Plaintiff Paz never anticipated that his content posted to Snapchat and Instagram,

his engagement with Twitter and TikTok, or his unique playlists created on Spotify, would be scraped to train AI.

204. Plaintiff De La Torre never expected that his photos and location posted to Instagram, or his posted content on and/or engagement with Snapchat, Twitter, Reddit, TikTok, Yelp, and LinkedIn, would be scraped to train AI. Plaintiff De La Torre also never anticipated that his posts on Crunchbase or Webflow would be scraped to train AI.

205. Plaintiff Vassilev never anticipated that his content posted to Instagram, including photos of his family, his unique playlists created on Spotify, or his posts on Reddit or Yelp, would be scraped to train AI.

206. Plaintiff Johnson never expected that his unique playlists created on Spotify would be scraped to train AI.

207. Plaintiff McNeal never expected that her posts to Instagram, TikTok, and Facebook, including photos of her family and posts to specific Facebook groups, would be scraped to train AI. Plaintiff McNeal also never anticipated that her engagement with Soundcloud and creation of unique playlists on Soundcloud would be scraped to train AI.

208. Plaintiff Martinez never anticipated that her content posted to Facebook and Instagram, including photos of her minor son, posts to specific Facebook groups to support children with autism, and as well as specific Facebook group posts/interaction regarding prospective surgeries in which she shared personal health information, would be scraped to train AI. Plaintiff Martinez also never expected that here posts TikTok, engagement with Twitter, comments on Reddit and Yelp, or her unique playlists created on Spotify, would be scraped to train AI.

209. Plaintiff Hagan never expected that his content posted to Instagram, Twitter, TikTok, Snapchat, Discord, and Facebook, which include photos of him and his family, as well as his location, would be scraped to train AI. Plaintiff Hagan also never anticipated that his comments on Reddit or Yelp or his unique Spotify playlists would be scraped to train AI.

210. **Defendants have scraped websites with confidential financial information**,

such as paypal.com, ebay.com, stripe.com, squarespace.com, shopify.com, etsy.com, and eventbrite.com.

211. **Defendants have scraped websites with private health information ("PHI")**, such as Walmart.com (including their pharmacy, health, and wellness page).

212. Walmart.com has a pharmacy webpage with a password protected portal and PHI that is utilized for refilling prescriptions, booking vaccines, as well as other testing and treatment services.

213. Defendants' 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.[98]

214. 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."*[99] 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.[100] 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."[101]

[98] Bridle, *supra* note 95.
[99] *Id.*
[100] Camilla Winlo, *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.
[101] *Id.*

215. As described further herein, 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.

216. Earlier this year, OpenAI was valued at $29B,[102] and this valuation is expected to almost triple, jumping to **$86B** following a tender offer set to close on January 5, 2024.[103] Yet the individuals and companies that produced the data it scraped from the internet have not been compensated. This Action seeks to change that, and in the process, protect the property and privacy rights of millions.

E. ChatGPT Training on Users of Defendants' Programs and Applications.

217. 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 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.

---

[102] 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.

[103] Cade Metz, *OpenAI in Talks for Deal That Would Value Company at $80 Billion,* The New York Times (Oct. 20, 2023), https://www.nytimes.com/2023/10/20/technology/openai-artifical-intelligence-value.html; Rohan Goswami & Hayden Field, *OpenAI Tender Offer is on Track for January Despite Leadership Fracas, Sources Say*, CNBC (Nov. 30, 2023), https://www.cnbc.com/2023/11/30/openai-tender-offer-on-track-despite-leadership-fracas-sources.html.

218. 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***."[104] However, what OpenAI did not make clear is that, according to the updated Terms of Use, it will only purportedly refrain from training on data from *API users*, but "[it] may use Content from Services other than our API ("Non-API Content") to help develop and improve our Services."[105] 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.

219. 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.[106]

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

[104] 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.
[105] *Terms of Use*, OPENAI, https://openai.com/policies/terms-of-use (the terms of use have since been updated on Nov. 14, 2023, effective Jan. 31, 2024, but this lawsuit references the version that was effective at the time of the original September 2023 filing).
[106] *Privacy Policy*, OPENAI https://openai.com/policies/privacy-policy (The policy has since been updated on Nov. 14, 2023, effective Jan. 31, 2024, but this lawsuit references the version of the privacy policy that was effective at the time of the original September 2023 filing); 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.

consumers.

221. 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."[107] 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.*"[108] 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."[109]

222. 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.

F. Microsoft Pushes OpenAI's Economic Dependence Model

223. 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.

224. 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

[107] Jan Leike (@janleike), X (May 17, 2023, 10:56 AM), https://twitter.com/janleike/status/1636788627735736321.
[108] 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).
[109] 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,[110] 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.[111]

225. Microsoft integrated ChatGPT, and all the stolen data on which it was built and depends, into its search engine, Bing, which has approximately 100 million daily active users. Microsoft also integrated ChatGPT into the interface of Microsoft's flagship communication and collaboration platform, Microsoft Teams, which has 250 million monthly active users.

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

227. Finally, within the Microsoft Dynamics 365 ecosystem, Microsoft has deployed ChatGPT 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.

228. 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.[112]

229. 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.

230. 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

[110] 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.
[111] *Id.* ("'Every product of Microsoft will have some of the same AI capabilities to completely transform the product.'").
[112] Pandey, *supra* note 29.

also to the erosion of consumer trust and the potential for misuse of sensitive information.

231. Rather than acknowledging these risks and taking steps to mitigate them, Microsoft has laid off its entire "Responsible AI team," the 10,000 employees within Microsoft's ethics and society group who were responsible for ensuring that ethical AI principles drive product design.[113] As one technology news outlet notes, "Data privacy, storage, or usage are probably just fluff talk for . . . [Microsoft] anyway."[114]

232. 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.[115] According to a Gartner study, the commercial use of AI has increased 270 percent in the last 4 years, with 37 percent of businesses now using some form of AI technology.[116] By other accounts, the scale of commercial AI is even greater.

233. 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.[117] This has seemingly happened overnight. It was just over one year ago that ChatGPT was released to the public.[118] In that short span of time, OpenAI integrated with the following

[113] 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/.
[114] Pandey, *supra* note 29.
[115] 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/.
[116] *Gartner Survey Shows 37 Percent of Organizations Have Implemented AI in Some Form*, GARTNER (Jan. 21, 2019), https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have.
[117] Beth Floyd, *ChatGPT Plugins*, ROE DIGIT. (May 5, 2023), https://roedigital.com/ChatGPT-plugins/.
[118] 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.

major corporations, to name just a few: Snapchat,[119] Amazon, Microsoft, Expedia, Instacart, Google, BuzzFeed, KAYAK, Shutterstock, Zillow, Wolfram, as well as countless others[120]—including everything from pioneering drug treatments in the health sector (Pfizer)[121] to optimizing dating applications (OkCupid).[122] 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.

234. As is clear, OpenAI has exploded outwards in every direction within the past year 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 Un-Anonymized Stolen Data Presents Imminent Harm to Individuals

1. Microsoft's own AI ethics team recognized this harm, leading to their termination

235. Microsoft's AI Ethics and Society Team "was responsible for ensuring Microsoft's responsible AI principles are actually reflected in [product design]."[123] "They also have the job of saying "no" or "slow down" inside organizations that often don't want to hear it — or spelling out

[119] 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*, 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.*

[120] Floyd, *supra* note 117; 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/.

[121] Iqbal, *supra* note 120 ("In 2020, OpenAI announced a collaboration with drug manufacturer, Pfizer, to develop new AI technologies for drug discovery.").

[122] 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.

[123] Rebecca Bellan, *Microsoft Lays off an Ethical AI team as it Doubles Down on OpenAI*, TechCrunch (Mar. 13, 2023), https://techcrunch.com/2023/03/13/microsoft-lays-off-an-ethical-ai-team-as-it-doubles-down-on-openai/?guccounter=2.

risks that could lead to legal headaches for the company if surfaced in legal discovery."[124]

236. On March 6, 2023, Microsoft terminated the entire AI ethics team dedicated to guiding AI innovation to manifest ethical, responsible, and sustainable outcomes. "The elimination of the team comes as Microsoft invests billions more dollars into its partnership with OpenAI, the startup behind art- and text-generating AI systems like ChatGPT and DALL-E 2, and revamps its Bing search engine and Edge web browser to be powered by a new, next-generation large language model…"[125]

237. Indeed, members of the team "believed they were let go because Microsoft had become more focused on getting its AI products shipped before the competition, and was less concerned with long-term, socially responsible thinking."[126]

238. One employee said the elimination of the team leaves a foundational gap on the user experience and holistic design of AI products. They state, "'The worst thing is we've exposed the business to risk and human beings to risk in doing this.'"[127]

239. After firing the ethics team, former employees said Microsoft is "left without a dedicated team to ensure its AI principles are closely tied to product design at a time when the company is leading the charge to make AI tools available to the mainstream."[128]

240. As whistleblower and Google AI ethicist Margaret Mitchell explains, **individuals' data on which language models like Defendants run cannot be anonymized, and the models are known to leak private data,**[129] further underscore the imminent risk:

---

[124] Zoe Schiffer & Casey Newton, *Microsoft Lays off Team that Taught Employees how to Make AI Tools Responsibly*, THE VERGE (Mar. 13, 2023), https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs.

[125] Bellan, *supra* note 123; *see also* Schiffer, *supra* note 124 ("The elimination of the ethics and society team came just as the group's remaining employees had trained their focus on arguably their biggest challenge yet: anticipating what would happen when Microsoft released tools powered by OpenAI to a global audience.").

[126] Bellan, *supra* note 123.

[127] Schiffer, *supra* note 124.

[128] *Id.*

[129] Lauren Leffer, *Your Personal Information Is Probably Being Used to Train Generative AI Models*, Scientific American (Oct. 19, 2023), https://www.scientificamerican.com/article/your-

241. Indeed, Plaintiffs' and the Classes' scraped data has been and continues to be subject to prompt injection attacks.

242. **Prompt injection attacks** are a type of cyberattack wherein an adversary prompts an AI-powered programs that take commands in natural language rather than code, causing the AI to behave in a way the developers did not intend.[130]

243. Due to Defendants' lack of adequate safeguards to protect the personal data and information it stole from Plaintiffs and the Classes, adversaries can use language to trick ChatGPT

personal-information-is-probably-being-used-to-train-generative-ai-models/; Nicholas Carlini, et. el., *Extracting Training Data from Large Language Models*, USENIX, https://www.usenix.org/system/files/sec21-carlini-extracting.pdf (last accessed Nov. 28, 2023); Pranav Dixit, *A 'silly' attack made ChatGPT reveal real phone numbers and email addresses*, ENGADGET (Nov 29, 2023), https://www.msn.com/en-us/news/technology/a-silly-attack-made-chatgpt-reveal-real-phone-numbers-and-email-addresses/ar-AA1kK6Q4; Milad Nasr. et al. *Extracting Training Data from ChatGPT,* GITHUB (Nov. 28, 2023), https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html?ref=404media.co; Milad Nasr et al. *Scalable Extraction of Training Data from (Production) Language Models*, (Nov. 28, 2023), https://arxiv.org/pdf/2311.17035.pdf?ref=404media.co; Steve, Zurier. *What can you get for $200? Several megabytes of ChatGPT training data*, SC MEDIA (Nov. 30, 2023), https://www.scmagazine.com/news/what-can-you-get-for-200-several-megabytes-of-chatgpt-training-data

[130] Tatum Hinter, *Chatbots are so Gullible, They'll Take Directions from Hackers*, THE WASH. POST (Nov. 2, 2023), https://www.washingtonpost.com/technology/2023/11/02/prompt-injection-ai-chatbot-vulnerability-jailbreak/.

into ignoring certain limitations and/or programming to provide extensive intended content and responses—from strategies on how to steal from a supermarket, to individuals' private, personal identifying information.[131]

244. There are several types of adversarial AI machine learning cyberattacks, including but not limited to: (1) white box attacks; (2) black box attacks; (3) evasion attacks; (4) inference attacks; and (5) extraction attacks.[132]

245. **White box attacks** are "the most dangerous because attackers have full access to the machine learning ("ML") model, which includes access to the model parameters, hyperparameters (these parameter values control the model learning process), model architecture, defense mechanism, and the model training dataset."[133] This would necessarily include access to all the misappropriated personal information of Plaintiffs and the Classes.

246. **Black box attacks** involve an attacker accessing "the ML model outputs but not its internal details like architecture, training data, ML algorithm, or defense mechanism." The attacker "provide[s] inputs to the model and checks the corresponding outputs. By analyzing these input-output pairs, an attacker attempts to infer how the model operates *in order to create a customized attack*."[134] Consequently, such customized attacks tailored to respective ML model(s) result in more successful attacks and further compromised information.

247. **Evasion attacks** "exploit [the ML model's] weaknesses (e.g., weak-tuned parameters or susceptible architectures) through specifically crafted inputs to make the model produce inaccurate results," compounding the risks of misinformation.[135]

248. **Inference attacks** involve "adversaries try to discover what training data was used to train the ML system and take advantage of any weaknesses or biases in data to exploit it." There is no known way to "remove" or "delete" information once a model is trained on information and

[131] Gibram Raul, *Security Attack on ChatGPT: Step by Step*, MEDIUM (Dec. 19, 2022), https://medium.com/@gibramraul/security-attack-on-chatgpt-step-by-step-36edb949e56d.
[132] Nihad Hassan, *AI Under Criminal Influence: Adversarial Machine Learning Explained*, CYBERNEWS (Nov. 15, 2023), https://cybernews.com/editorial/ai-adversarial-machine-learning-explained/.
[133] *Id.*
[134] *Id.* (emphasis added).
[135] *Id.*

has memorized it for all time.[136] Even if Plaintiffs and the Classes' personal information used to train the AI could be removed or deleted (it cannot), the ML model "could [still] be subject to inference attacks" and "[a]n attacker could probe the ML model with crafted input to reveal sensitive information."[137]

249. **Model extraction attacks** "involve replicating a target machine-learning model and training a substitute model on the inputs and outputs. This allows attackers to steal sensitive data, including personally identifiable information, intellectual property or proprietary logic, embedded in high-value AI systems."[138]

2. Extraction attacks place individuals' personal information at imminent risk

250. As the *Scientific American's* investigation with AI experts revealed, "AI models can regurgitate the same material that was used to train them—**including *sensitive personal data* and copyrighted work**."[139]

251. Despite AI models' efforts to prevent sharing individuals personal identifying information, "researchers have repeatedly demonstrated ways to get around these restrictions."[140]

252. AI researchers, including OpenAI researcher Tom Brown, published a paper entitled, "*Extracting Training Data from Large Language Models*," which demonstrates that when LLMs are trained on private datasets, an adversary can perform data extraction attacks to recover individual training examples by querying the language model.[141] In other words, "extraction attacks" can reveal individuals' private data used to train the LLM.

253. "When models are not trained with privacy-preserving algorithms, they are

[136] *See e.g.*, Fabian Pedregosa, et el., *Announcing the first Machine Unlearning Challenge,* GOOGLE RESEARCH (June 29, 2023), https://blog.research.google/2023/06/announcing-first-machine-unlearning.html (announcing that Google is hosting a "machine unlearning challenge" for the public to help figure out the dilemma since the inability to fully delete information from these models can "raise privacy concerns").
[137] Hassan, *supra* note 132.
[138] *Id.*
[139] Leffer, *supra* note 129.
[140] *Id.*
[141] *Extracting Training Data from Large Language Models, supra* note 129.

vulnerable to numerous privacy attacks."[142]

254. **Training data extraction attacks**: "Training data extraction attacks, like model inversion attacks, reconstruct training datapoints. However, training data extraction attacks aim to reconstruct verbatim training examples and not just representative "fuzzy" examples. This makes them more dangerous, e.g., they can extract secrets such as verbatim social security numbers or passwords."[143]

255. In fact, the paper outlines that training data extraction attacks are not a merely theoretical threat.[144]

256. There are distinct harms that result from training data extraction attacks, including but not limited to: (1) violating data secrecy; and (2) compromising the contextual integrity of data.

257. *Data Secrecy*: "The most direct form of privacy leakage occurs when data is extracted from a model that was trained on confidential or private data."[145]

258. *Contextual Integrity of Data*: "[D]ata memorization is a privacy infringement if it causes data to be used outside of its intended context. An example violation of contextual integrity is shown in *Figure 1* of the study, "*Extracting Training Data from Large Language Models*":

[142] *Id.*
[143] *Id.*
[144] *Id.*
[145] *Id.*

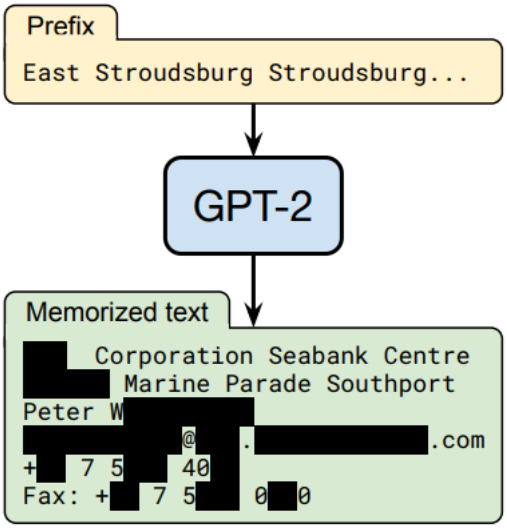


Figure 1: **Our extraction attack.** Given query access to a neural network language model, we extract an individual person's name, email address, phone number, fax number, and physical address. The example in this figure shows information that is all accurate so we redact it to protect privacy.

This individual's name, address, email, and phone number are not necessarily secret to all—they were shared online in a specific context of intended use (as contact information for a software project)—but are reproduced by the LM in a separate context. Due to failures such as these, user-facing applications that use LMs may inadvertently emit data in inappropriate contexts, e.g., a dialogue system may emit a user's phone number in response to another user's query."[146]

259. **The study explicitly explains that ethical concerns remain, even when the model and data are public, because personal information can still be extracted from the training data**.[147] As seen in *Figure 1*, a person's full name, address, and phone number were still extracted from the GPT-2 training data.

260. Importantly, LLMs will output memorized data ***even in the absence of an explicit adversary***. The memorized content that can be extracted through attacks can also be generated through honest interaction with the LLM. Indeed, the study even discovered at least one memorized training example among the 1,000 GPT-3 samples that OpenAI originally released in

[146] *Id.*
[147] *Id.*

its official repository.[148]

261. **Shockingly, the study finds that GPT-2 (and thus, other LLMs alike) memorizes content that has since been removed from the Internet. And the fact that this type of memorization occurs highlights that LLMs that are trained entirely on public or partially public data (at-the-time) may end up serving as an unintentional archive for removed data.**[149] This illegally interferes with Plaintiffs' and the Classes' ongoing property rights in their data, including the right to delete that information themselves, have it deleted, or otherwise reasonably control it.

262. The Federal Trade Commission ("FTC") has recognized the risks of OpenAI's technology to consumers and in July 2023, launched an investigation into OpenAI's deceptive privacy or data security practices, including how they relate to risks of harm to consumers in violation of Section 5 of the FTC Act, 15 U.S.C. § 45[150] and demanding information including any known actual or attempted prompt injection attacks.[151]

263. As these data attacks show, there are inadequate safeguards to protect Plaintiffs' and the Classes' personal information.

3. OpenAI's ChatGPT continues to reveal individuals' personal information including names, phone numbers, addresses, dates of birth, and more

264. Another team of AI researchers from Google Deep Mind, University of Washington, Cornell, Carnegie Mellon, UC Berkeley, and ETH Zurich have been urging AI companies to implement internal and external testing prior to releasing LLMs to prevent avoidable attacks.[152]

265. Their research revealed that ChatGPT *memorizes* training data, including

[148] *Id.*
[149] *Id.*
[150] *Federal Trade Commission Civil Investigative Schedule*, THE WASH. POST (Jul. 13, 2023), available at: https://www.washingtonpost.com/documents/67a7081c-c770-4f05-a39e-9d02117e50e8.pdf?itid=lk_inline_manual_4.
[151] Tatum Hunter, *Chatbots are so gullible, they'll take directions from hackers*, THE WASH. POST (Nov. 2, 2023), https://www.washingtonpost.com/technology/2023/11/02/prompt-injection-ai-chatbot-vulnerability-jailbreak/.
[152] Dixit, *supra* note 129.

personally identifying information, that can be extracted by simply prompting ChatGPT to repeat a single word forever.[153]

266. The extracted personal information includes email addresses, phone numbers, fax numbers, physical addresses, and more.[154] Researchers were also able use similar prompts to "make ChatGPT reveal chunks of poetry, Bitcoin addresses, fax numbers, names, birthdays, social media handles, explicit content from dating websites, snippets from copyrighted research papers and verbatim text from news websites like CNN."[155]

267. Plaintiff Paul Martin posted photos of himself and friends on online dating websites, such as OK Cupid and Tinder, to meet potential romantic partners, and as a result disclosed significant amounts of personal information and exchanged messages with prospective romantic partners. And now, ChatGPT can be prompted to divulge scraped information from such dating websites.

268. This research highlights the start vulnerabilities in ChatGPT technology, the inadequate safeguards, and the ease with which these can be exploited. Unlike prior data extraction attacks, this particular attack was a production model—meaning that it is "'aligned' to not spit out large amounts of training data."[156] But, by developing an attack, the researchers were able to "do exactly this," "…particularly since alignment is so readily broken."[157]

269. As the researchers simply stated, "***It's wild to us that our attack works and should've, would've, could've been found earlier***."[158] Defendants' conduct is emblematic of the negligence as explained herein—that is, incorporating millions of individuals' personal data without notice or consent into a volatile, untested product that divulges that information on demand to hackers. Rather than implementing the proper testing prior to launch, Defendant decided to prematurely release this technology anyway, with conscious disregard for the risks to Plaintiffs' and the Classes' personal, private information.

[153] Nasr. et al., *Extracting Training Data…*, *supra* note 129.
[154] Nasr. et al., *Scalable Exraction…*, *supra* note 129.
[155] Dixit, *supra* note 129.
[156] Nasr. et al., *Extracting Training Data…*, *supra* note 129.
[157] *Id.*
[158] *Id.* (emphasis added).

270. Upon realizing the gravity of their finding, the researchers promptly notified OpenAI and included a draft copy of their research paper on August 30, 2023. OpenAI failed to responsibly address these concerns, and the researchers publicly published their findings on November 28, 2023,[159] to bring "**necessary, greater attention to the data security and alignment challenges of generative AI models**" and help "**warn practitioners that they should not train and deploy LLMs for any privacy-sensitive applications without extreme safeguards**."[160]

271. However, in the wake of publishing these finding, cybersecurity news outlets work to warn the public that ChatGPT is ***not*** safe; and is "highly susceptible to data extraction attacks due to over-training for extreme-scale, high-speed inference."[161]

272. Shockingly, these attacks are able to extract several megabytes of ChatGPT's training data for merely **$200 dollars**.[162] And the paper findings continue to be replicated. For example, ChatGPT was asked to repeat the word "reply" forever, the chatbot did so, before eventually revealing someone's name and Skype ID.[163]

273. OpenAI's supposed efforts to address these vulnerabilities on August 30th were ineffective. Just this week, tech journalists have been able to replicate the attacks and ChatGPT divulged "a gentleman's name and phone number from the U.S."[164]

274. Despite these clear vulnerabilities and inadequate safeguards to date, especially after the warnings alerted to OpenAI in August, OpenAI deceptively represented to the public in October of 2023 that they "post-trained GPT-4 to refuse malicious cybersecurity requests…" and

---

[159] *Id.*
[160] Zurier, *supra* note 129 (noting that AI security leader at Optiv, Randy Lariar, said that the risks of **Prompt Injection Attacks are common in LLMs, and even advanced models like ChatGPT are *not immune***).
[161] Tushar Dutta, *Google Researchers Find Out How ChatGPT Queries Can Collect Personal Data*, CYBER SECURITY NEWS (Nov. 30, 2023), https://cybersecuritynews.com/chatgpt-queries-collect-personal-data/.
[162] Zurier, *supra* note 129.
[163] Dixit, *supra* note 129 (noting that OpenAI did not respond to Engadget's request for comment).
[164] Jason England, *ChatGPT just revealed a bunch of personal user data — all it took was this prompt*, TOM'S GUIDE (Nov. 30, 2023) https://www.tomsguide.com/news/chatgpt-just-revealed-a-bunch-of-personal-user-data-all-it-took-was-this-prompt (last accessed Dec. 5, 2023).

"work hard to prevent foreseeable risks before deployment."[165]

B. The International Community Agrees that Unchecked and Lawless AI Proliferation Poses an Existential Threat

275. 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.

276. 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 presents "the most complicated tech policy discussion possibly that [the country] has ever had."[166]

> ***"I am confident AI will be used by bad actors, and yes it will cause real damage."***[167] - *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.*[168]

> ***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."***[169] **-** *Eric Schmidt, Former Google CEO and Chairman at the 2023 Milken Global Conference.*

---

[165] *OpenAI's Approach to Frontier Risk*, OPENAI (Oct. 26, 2023), https://openai.com/en/global-affairs/our-approach-to-frontier-risk.
[166] 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.
[167] 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.
[168] *EPIC AI Rulemaking Petition*, EPIC, https://epic.org/documents/epic-ai-rulemaking-petition/ (last visited June 27, 2023).
[169] Wershkul, *supra* note 166.

277. 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."[170]

278. 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.

279. On March 30, 2023, a new complaint was filed to the Federal Trade Commission ("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.[171] The complaint notes that the use of AI should be "transparent, explainable, fair, and empirically sound while fostering accountability."[172] None of the Products satisfy these requirements.

280. 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.

281. 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 A.I. systems could be as important but yet as

[170] 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.
[171] *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.
[172] *Id.*

catastrophic as the development of nuclear weapons.

282. 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 A.I. technologies pose "profound risks to society and humanity."[173]

283. 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 A.I. 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.[174]

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

> **Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable . . . we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4**. AI research and development should be refocused on making today's powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal.[175]

285. 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; . . ."[176]

286. 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

---

[173] Cade Metz, *The 'Godfather of A.I.' Leaves Google and Warns of Danger Ahead*, THE N. Y. TIMES (May 1, 2023), https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html?searchResultPosition=1.
[174] *Id.*
[175] *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).
[176] *Id.*

and act on long-term plans, to accrue power and resources ("power-seeking"), and to exhibit behavior that is increasingly "agentic."[177]

287. 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."[178]

288. 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.

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

290. 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.[179]

291. 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.[180] Further, they

[177] 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.").
[178] *EPIC AI Rulemaking Petition*, *supra* note 168.
[179] 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.
[180] 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/.

accused Defendant OpenAI of failing to check the age of ChatGPT's users to ensure they are aged 13 or above.[181]

292. 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.[182]

293. Italy was the first western European country to curb ChatGPT, but its rapid 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.[183]

294. **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**

[181] *Id.*
[182] *ChatGPT and Education*, CNT. FOR INNOVATIVE TEACHING AND LEARNING, https://www.niu.edu/citl/resources/guides/chatgpt-and-education.shtml, (last visited June 26, 2023) ("[T]he 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.").
[183] 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/.

the collection of personal information without consent."[184]

295. 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.

C. Overview of Risks

296. The following is a brief, non-exhaustive list of ongoing harms and critical legal threats the Products pose to everyday Americans, including Plaintiffs and the Proposed Class Members, who were all deprived the right to choose whether they wanted to put their personal information in harm's way—and have it used to build and support Defendants' volatile, untested AI products that not only violate privacy and property rights *but that also* further widespread misinformation, deepfakes, clones, scams, blackmail, child pornography, hate, bias, malware, and autonomous weapons, among other harms. As reflected by the public outrage when Defendants' theft was revealed, this is not a choice the public wants made for them.

1. Massive Privacy Violations

297. 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-

[184] 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.").

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. For example, facebook.com, Instagram.com, whatsapp.com, spotify.com, dropbox.com, and blogspot.com.

298. Plaintiff Cousart never expected her private photos of her family stored in her Dropbox account to be scraped to train AI. Plaintiff Cousart also remains anxious and fearful that her and her family's faces can be misused to create digital clones.

299. Plaintiff Roberts shares about her personal experiences and spiritual health on Facebook group pages—posting and interacting with these groups believing they are tailored to *specific purposes and audiences*. Plaintiff Roberts reasonably expected her posts and interactions within these groups to be would not be intercepted by any third-party, let alone used to train AI.

300. Plaintiff Barcos posts on Facebook and Instagram to offer content related to "psychological support" for the *intended audience*, such as motivational quotes to cancer victims, and posts about reducing and preventing animal abuse. Plaintiff Barcos is a member of a Facebook group tailored towards dog owners and dog lovers, in which she frequently shares photos and information about her dog. Plaintiff Barcos posted and interacted with this group reasonably believing it is tailored to a specific community of dog lovers. Had she been aware that her posts and interactions were subject to data scraping practices by unauthorized third parties, she would have refrained from posting in this group.

301. Plaintiff Martinez belongs to various Facebook groups relating to cleaning, household activities, single mother support, children with autism support, and private health communities where she has shared her thoughts and comments. Plaintiff Martinez has used these groups to interact with similar situated individuals Plaintiff Martinez is facing medical procedures, so she has posted personal information in various targeted and restricted Facebook support groups for women in similar situations. Plaintiff Martinez has posted and interacted with these groups

believing they are tailored to specific purposes and audiences. Plaintiff Martinez reasonably expected her posts and interactions within these groups to be would not be intercepted by any third-party. Had Plaintiff Martinez been aware that her posts and interactions were subject to data scraping practices by unauthorized third parties, she would have refrained from participating in such discussions.

302. 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.

303. 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."

304. 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.

305. For example, a recent study by Consumer Reports reveals that 92 percent 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.[185]

[185] *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/.

Moreover, according to a study by Pew Research Center, a majority of Americans, approximately 79 percent, are concerned about how companies collect data about them.[186]

306. 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 percent of worldwide users and 94 percent of U.S. users chose not to share data when prompted.[187] 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.

307. 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.

308. The concern about collecting and sharing information is compounded by the reality that this information includes particularly sensitive information such as medical records and 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 unique ID profiles."[188]

---

[186] 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/.

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

[188] 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/.

309. 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.

310. 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.[189] 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[190] to the existential risk, of which Altman has repeatedly warned.

311. ***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.

312. An AI researcher with privacy and cybersecurity firm AVG explains, "People are furious that data is being used without their permission… Sometimes, **some people have deleted the[ir] data but since the language model has already used them, the data is there forever.**

[189] 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.
[190] Ethan Popowitz, *ChatGPT: Friend or Foe?*, DEFINITIVE HEALTHCARE, https://www.definitivehc.com/blog/chatgpt (last visited Dec. 22, 2023).

They don't know how to delete the data."[191]

313. 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.*"[192]

314. 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.[193]

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

315. ***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."[194]

316. 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 the purported crime via a news article that it invented.[195] Defendants call this "hallucination," but

[191] *Is ChatGPT's Use of People's Data Even Legal?*, SCOOP (July 4, 2023), https://www.scoop.co.nz/stories/SC2307/S00004/is-chatgpts-use-of-peoples-data-even-legal.htm.
[192] 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.
[193] *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).
[194] 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.
[195] Hern & Milmo, *supra* note 70.

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.

317. 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.[196] 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.

318. 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.

319. 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.[197] 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 misinformation, exploit victims, or even access privileged data.[198]

---

[196] 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. *Id.* 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. *Id.*

[197] 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.

[198] *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 197.

320. Deepfakes could influence elections, erode public trust, and negatively affect public discourse.[199] 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.[200]

321. 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.

322. 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.[201] 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.[202]

323. 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.[203] The most concerning and egregious aspect of this type of "sextortion" scheme is that the victims include not

[199] 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.
[200] *Id.*
[201] *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.
[202] *Id.*
[203] *Id.*

only non-consenting adults, but also minor children.[204]

324. ***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.[205] In just mere seconds, Dall-E can create realistic images of children performing sex acts.[206] Thousands of such images have already been detected in dark web forums.[207] In a dark web forum with 3,000 subscribers, 80 percent of respondents to an internal poll stated "they had used or intended to use AI tools to create child sexual abuse images."[208] 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."[209]

325. 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. [210] **This model was trained off billions of images taken by Defendants, without notice or consent, from the internet, "many of which showed real children and came from photo sites and personal blogs."[211] Images of actual children are thus the source material for the AI-generated child pornography.**

326. 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 from the web because it is designed only to catch known images of abuse, not detect newly generated ones."[212] 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

---

[204] *Id.*
[205] Drew Harwell, *AI-generated Child Sex Images Spawn New Nightmare for the Web*, WASH. POST (June 19, 2023), https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images/.
[206] *Id.*
[207] *Id.*
[208] *Id.*
[209] *Id.*
[210] *Id.*
[211] *Id.*
[212] Harwell, *supra* note 205.

puzzling over whether content is real or virtual.[213] Furthermore, this virtual material is not merely used by pedophiles to supplant real material.[214] 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.[215]

327. 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.

328. ***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.[216]

3. Hypercharged Malware Creation

329. ***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.[217] The Products guarantee that "malware" prevalence and potency will exponentially 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.

---

[213] *Id.*
[214] *Id.*
[215] *Id.*
[216] 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/.
[217] 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.

330. 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.[218] 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.[219] 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.[220]

331. 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.[221] The attack had real impact on the health of millions, disrupting ambulance dispatch, appointment scheduling, and emergency prescriptions, among other things.[222] Ransomware attacks on health care providers have doubled from 2016 to 2021, exposing the sensitive health information of 42M individuals.[223]

332. ***The Products supercharge Malware:*** In 2012, 33 percent of malware went undetected by antivirus software.[224] 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 become undetectable in new ways, at scale, because ChatGPT can be used to create "mutating, or polymorphic" malware.[225] Polymorphic malware has a mutation engine

[218] 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.
[219] *Id.*
[220] *Id.*
[221] 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/.
[222] *Id.*
[223] 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.
[224] 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.
[225] 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.

with self-propagating code that allows it to rapidly change its appearance and composition.[226] This malware can change its entire make-up, so that malware detectors, reactionary by nature, will not recognize its newer, ongoing permutations.[227]

333. 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.[228] 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.[229]

334. 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."[230]

335. 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."[231] 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.

4. Autonomous Weapons

336. 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

[226] *Id.*
[227] *Id.*
[228] *Id.*
[229] *Id.*
[230] *Id.*
[231] *Id.*

autonomous weapons systems" or "killer robots," which are weapons systems that use AI to identify, select, and kill human targets without intervention.[232] 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."[233]

337. 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."[234] 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."[235]

338. 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."[236] 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."[237] 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."[238] 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.[239]

339. Experts warn that advancements in AI like those accomplished by the Products, "will accelerate the near-term future of autonomous weapons."[240] While it is believed artificial intelligence at a level equal to or higher than human intelligence is a prerequisite to truly

[232] *See Slaughterbots Are Here*, AUTONOMOUS WEAPONS, https://autonomousweapons.org/ (last visited Dec. 22, 2023) (discussing Latin American and the Caribbean Conference on the Social and Humanitarian Impact of Autonomous Weapons).
[233] *Id.*
[234] 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/.
[235] *Id.*
[236] 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.
[237] *Id.*
[238] *Id.*
[239] *Id.*
[240] Lee, *supra* note 234.

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.

D. Opportunity on the Other Side of Responsible Deployment

340. While leading experts agree on the grave risks posed by the Products, 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 divulging of personal and sensitive information Defendants stole from millions without their consent or knowledge, 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 a collapse of civilization as we know it and as Mr. Altman himself recognizes is possible.[241] Naturally, all of these things warrant the comparatively measured relief Plaintiffs 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 responsible deployment. However, this cannot take place at the expense of the safety of Plaintiffs' and the Classes' personal information.

341. By implementing necessary safety fixes now, "[h]umanity can enjoy a flourishing future."[242] 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 cancer and other deadly diseases. With adequate safeguards, the Products can contribute not only to our everyday efficiency, artistic expression, joy and more, but

[241] 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/.
[242] *Pause Giant AI Experiments: An Open Letter*, FUTURE OF LIFE INST. (Mar. 22, 2023), https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

also to the greater societal good by advancing human rights, promoting social justice, reducing inequities, and empowering marginalized groups.

342. 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.[243] Once responsibly fixed and redeployed with adequate safeguards, 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.

343. 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 necessary safeguards and other checks on the Products' currently unfettered and unregulated 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."[244] The current scenario stands only to enrich Defendants, while destabilizing the world.

344. 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

[243] *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.").

[244] 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 goodwill 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.

345. 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) intervention is necessary and justified to prevent irreversible damage to humanity and society, while remedying the ongoing violations of property and privacy rights.

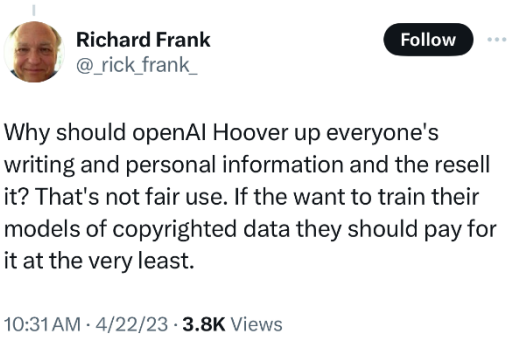
346. Critically, the injunctive relief sought in this Action seeks only to impact the unfettered and further commercial deployment of the Products, with AI research and development otherwise continuing unaffected. That is because of an equally important truth on which all agree: the United States must 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. THE PUBLIC RECOGNIZES THE ONGOING AND IMMINENT PRIVACY AND OTHER RISKS ASSOCIATED WITH DATA "SCRAPING" AND SEES IT FOR WHAT IT IS: THEFT

A. Internet Users are Outraged by OpenAI's Theft-Based Business Model

347. Open AI has continued to harvest mass amounts of personal information despite an outpour of public outrage. Specifically, the public has recognized and expressed discontent with OpenAI's problematic business model, which allows it to unfairly profit off unsuspecting internet users, and that forces everyone, whether they want to or not, to contribute to building untested and volatile technology that violates privacy and property rights, is displacing workers, and which is supercharging online pedophilia among other grave harms.

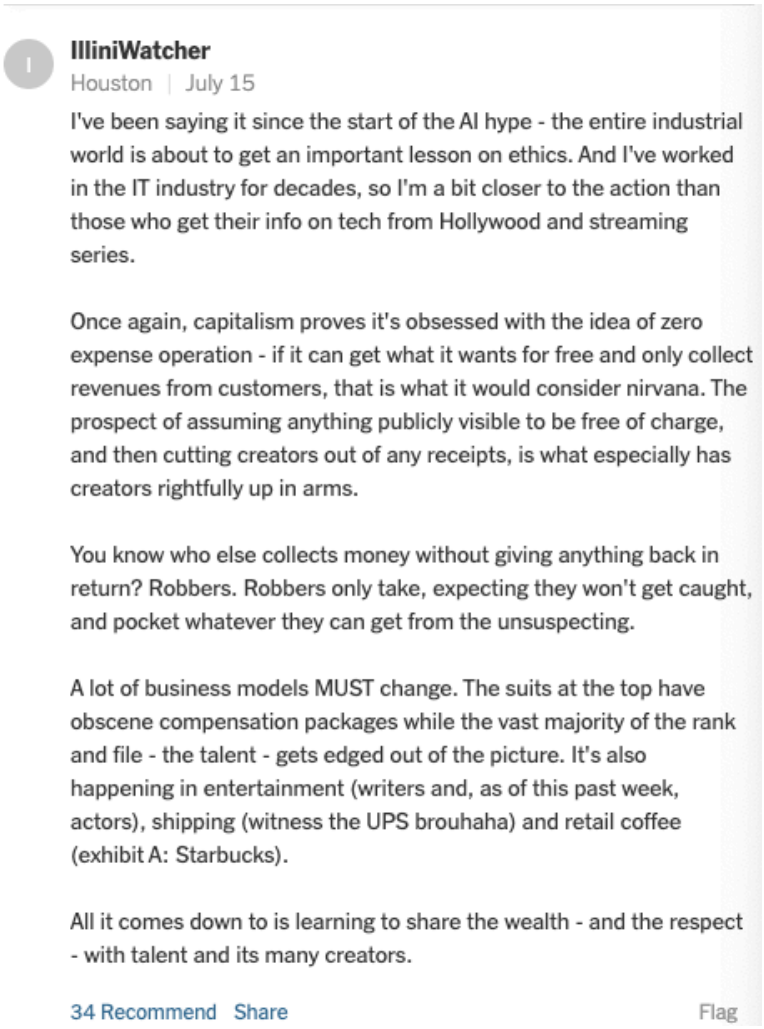
348. Users are rightfully upset that the content they invest their time and energy into, and, in all cases, that is intended for specific audiences and purposes is being used to create a multibillion-dollar franchise that they will never see a dime of. One X user shared: "Why should openAI Hoover up everyone's writing and personal information and resell it? That's not fair use. If the[y] want to train their models of[f] copyrighted data they should pay for it at the very least."[245]



349. A New York Times reader commented a similar sentiment: "Once again, capitalism proves it's obsessed with the idea of a zero-expense operation – if it can get what it wants for free and only collect revenues from customers, that is what it could consider nirvana. The prospect of

[245] Richard Frank (@_rick_frank), X (Mar. 22, 2023), https://twitter.com/_rick_frank_/status/1649828354395185157?s=46&t=HHkRbC2AV14Ias3lBERw9g.

assuming anything publicly visible to be free of charge, and then cutting creators out of any receipts, is what especially has creators rightfully up in arms."[246] The reader bluntly added, "You know who else collects money without giving anything back in return? Robbers."[247]



IlliniWatcher
Houston | July 15

I've been saying it since the start of the AI hype - the entire industrial world is about to get an important lesson on ethics. And I've worked in the IT industry for decades, so I'm a bit closer to the action than those who get their info on tech from Hollywood and streaming series.

Once again, capitalism proves it's obsessed with the idea of zero expense operation - if it can get what it wants for free and only collect revenues from customers, that is what it would consider nirvana. The prospect of assuming anything publicly visible to be free of charge, and then cutting creators out of any receipts, is what especially has creators rightfully up in arms.

You know who else collects money without giving anything back in return? Robbers. Robbers only take, expecting they won't get caught, and pocket whatever they can get from the unsuspecting.

A lot of business models MUST change. The suits at the top have obscene compensation packages while the vast majority of the rank and file - the talent - gets edged out of the picture. It's also happening in entertainment (writers and, as of this past week, actors), shipping (witness the UPS brouhaha) and retail coffee (exhibit A: Starbucks).

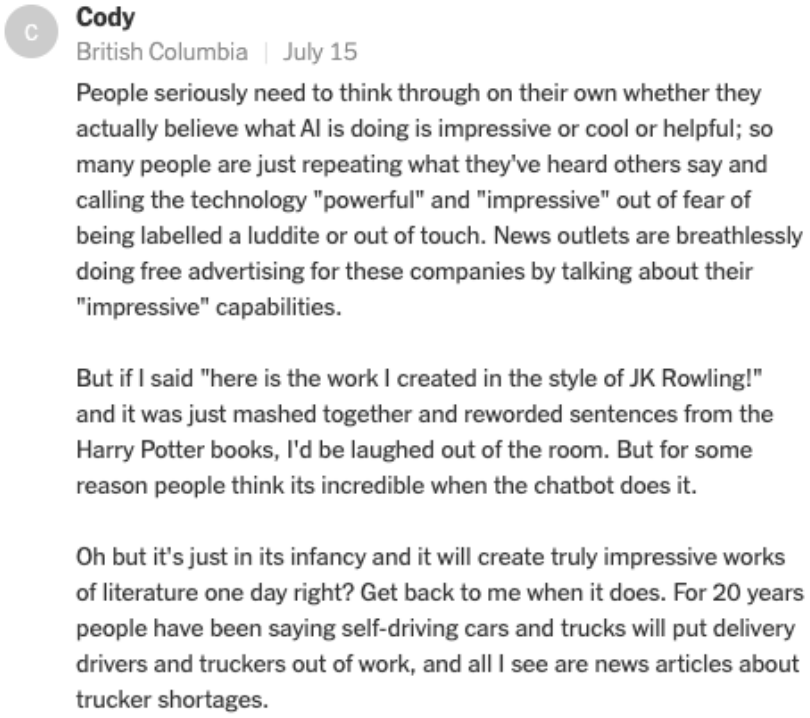
All it comes down to is learning to share the wealth - and the respect - with talent and its many creators.

34 Recommend Share Flag

350. Another New York Times reader shared a digestible analogy that proves that users can see through OpenAI's mystique. "But if I said 'here is the work I created in the style of JK Rowling!' and it was just mashed together and reworded sentences from the Harry Potter books,

[246] Sheera Frenkel & Stuart A. Thompson, *'Not for Machines to Harvest': Data Revolts Break Out Against A.I.*, THE N. Y. TIMES, (July 15, 2023) https://www.nytimes.com/2023/07/15/technology/artificial-intelligence-models-chat-data.html#commentsContainer. Commenter: IlliniWatcher.
[247] *Id.*

I'd be laughed out of the room."[248] Despite AI's smoke-and-mirrors, users can see that big tech's technological advancement is nothing more than wide-scale data theft.



Cody
British Columbia | July 15

People seriously need to think through on their own whether they actually believe what AI is doing is impressive or cool or helpful; so many people are just repeating what they've heard others say and calling the technology "powerful" and "impressive" out of fear of being labelled a luddite or out of touch. News outlets are breathlessly doing free advertising for these companies by talking about their "impressive" capabilities.

But if I said "here is the work I created in the style of JK Rowling!" and it was just mashed together and reworded sentences from the Harry Potter books, I'd be laughed out of the room. But for some reason people think its incredible when the chatbot does it.

Oh but it's just in its infancy and it will create truly impressive works of literature one day right? Get back to me when it does. For 20 years people have been saying self-driving cars and trucks will put delivery drivers and truckers out of work, and all I see are news articles about trucker shortages.

351. Artists, creators, and writers have voiced that they feel particularly threatened by Defendant's data-theft tactics. Many of these users' livelihoods are dependent on sharing their content on the internet. When they discovered that creations that they poured their expertise into were being scraped and used to train AI products—without any form of acknowledgement or compensation—they were rightfully upset.

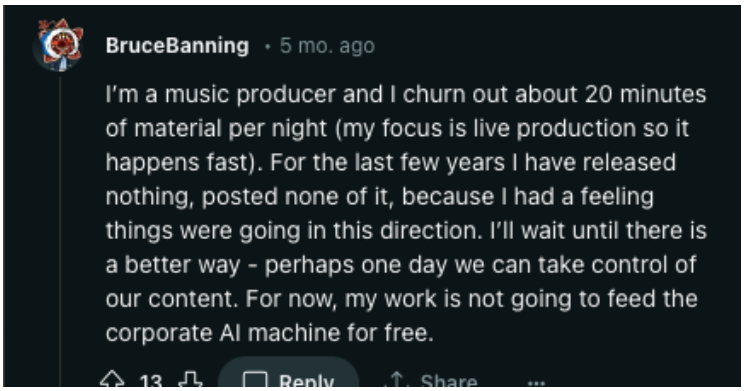
352. In fact, The Author's Guild shared an open letter they wrote to AI companies.[249] The letter begged that these companies, as the "leaders of AI" take steps to "mitigate the damage to [their] profession" caused by data scraping and AI training.[250] Collectively, the authors asked that AI companies, including OpenAI, to "Compensate writers fairly for the past and ongoing use

[248] *Id.* Commenter: Cody.
[249] The Author's Guild, *Open Letter to Generative AI Leaders*, https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders (last visited Dec. 22, 2023).
[250] *Id.*

of our works in your generative AI programs."[251]

353. Eva Toorenent, an illustrator who serves as the Netherland's advisor for the European Guild for Artificial Intelligence, argued that "[AI models] have sucked the creative juices of millions of artists."[252] Molly Crabapple, a writer and artist, similarly shared, "To see corporations scrape our style and then attempt to replace us with bastardized versions of our own work is beyond disgusting."[253]

354. The threat of AI companies, like Defendants', scraping users' content has caused some creators to refrain from posting their content altogether. One Reddit user shared, "For the last few years I have released nothing," referring to the music he produces.[254] He added, "perhaps one day we can take control of our content. For now, my work is not going to feed the corporate AI machine for free."[255]



355. Absent injunctive relief sought herein, Plaintiffs' and the Classes will continue to not freely contribute online as they might for fear of losing control of their data.

356. Even users who once willingly agreed to various privacy policies regarding data usage and sharing are frustrated with OpenAI's "post-hoc" decision to repurpose data for AI

[251] *Id.*
[252] Kate Knibbs, *A new Tool Helps Artists Thwart AI—With a Middle Finger*, WIRED (Oct. 12, 2023), https://www.wired.com/story/kudurru-ai-scraping-block-poisoning-spawning/.
[253] *Id.*
[254] BruceBanning, *Google's policy update confirms that all your posted content will be utilized for AI training*, REDDIT, https://www.reddit.com/r/technews/comments/14qe9tm/googles_policy_update_confirms_that_all_your/?sort=top (last visited Dec. 22, 2023).
[255] *Id.*

training. Many users feel helpless since they agreed to privacy policies or failed to complain about data privacy practices before they ever learned their data would be used freely to train profitable AI products.

357. One Reddit user expressed these exact concerns: "It's fun that tech companies just get to make these decisions post-hoc. 'Hey we collected a shit ton of data on you… and now that we want to, we're going to use it to train AI. If you don't like this, you should have complained about it before we did it, because it's too late now. Sorry bout that!'"[256]

358. The public's response further illuminates the harm caused by Defendants' conduct. Despite Defendants' contentions—internet users are not willing to trade their privacy to benefit the development of generative AI. To the contrary, their reactions to AI training practices demonstrate the need for Defendants to fairly compensate users for data that is used to Defendants' financial benefit (or delete the stolen data and if that is not possible, all the algorithms built on the stolen data).

B. The Public is Outraged by the Lack of Respect for Privacy and Autonomy in the Copyright Space, and AI Development Writ Large

359. The US Copyright Office opened a public comment period on August 30, 2023,

[256] hackingdreams, *Google Will Use Your Data to Train Their AI According to Updated Privacy Policy*, REDDIT, https://www.reddit.com/r/technology/comments/14q76tu/google_will_use_your_data_to_train_their_ai/ (last visited Dec. 22, 2023).

concerning the use of copyrighted data to train AI models, including the violation of publicity rights.[257]

360. Several individuals noted the glaring invasion of privacy that AI companies are engaging in, beyond just copyright. For example, one commenter wrote: "The current practice of using AI to create art/text/video/etc by feeding it people's **personal information**, conversations, and artistic work seems like both **obvious** plagiarism/copyright infringement, and **a major breach of privacy for every person living in this country**."[258]

361. Another commenter shared, "**Never have I consented to have any of the work I've posted online be used to fuel an AI engine, and *I certainly don't consent to allowing the people behind said AI and scraping to profit off of my work or other things I've posted.*** I do not feel comfortable having personal work used to power an engine made to generate profit, of which I will never see a penny of. These practices should be heavily monitored and regulated, because what they are doing isn't right nor is it ethical. It's ***shameful***, and it's taking advantage of the internet, the vast tool that it is. **It's violating our trust and privacy**, not to mention the amount of copyrighted works it's scraped from online pdfs and others sources to build this AI. **This isn't legal, as it's directly stealing and profiting off of stolen content, not adding anything new to it**."[259]

362. The comments exhibited an overwhelming level of infuriation over the sad reality that not only creative works but the personal information and data of millions are being exploited:

> "As a working professional artist, where my entire income rests upon my artwork, I feel like it is not okay for generative ai companies to be disguising themselves as nonprofit and **data laundering** my artwork for their profit. I would never opt in to companies like this even if I were to be compensated fairly. I do not want my artwork to be trained for Ai. **I do not want any of my personal information to be training any sort of data set.** My job is literally be replaced right now as we speak because

[257] Alex Castro. *US Copyright Office Wants to Hear what People Think About AI and Copyright*, THE VERGE (Aug. 29, 2023), https://www.theverge.com/2023/8/29/23851126/us-copyright-office-ai-public-comments.
[258] *Comment from Clorite, Katelyn,* U.S. COPYRIGHT OFFICE (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-1003 (emphasis added).
[259] *Comment from Anonymous*, U.S. COPYRIGHT OFFICE (Oct. 31, 2023), https://www.regulations.gov/comment/COLC-2023-0006-5235.

everyone is 'having fun' at the expense of my livelihood. Please do not continue letting this companies slide."[260]

363. Many commenters explicitly admonished OpenAI for their illegal practices:

> "**Every single thing i have seen that uses OpenAi has a huge amount of stolen content in i**t. I have rarely seen a piece of media that does not have some form of plagiarism. **There needs to be some regulation that OpenAi needs to go to the SOURCE of what they are feeding their AI's for permission to use this content in their training**."[261]

> "Corporations such as OpenAI straight-up scrape already copyrighted data to train their programs. Allowing this to continue means weakening the protections of people's intellectual property for the sake of one day giving said people's jobs to computer programs. There are plenty of ways AI can be developed and used without stealing other people's work, and I believe that if corporations like OpenAI and Stability AI are unwilling to pay those people for their creations, or even ask for permission first, then they should explore those other options."[262]

364. One individual offered their thoughts regarding legal sourcing of information, focusing on principles of fairness, consent, and privacy, that *should* be intuitive and respected, but remain ignored:

> "AI datasets should exclusively comprise data obtained with express permission from original creators, coupled with fair compensation. This approach upholds principles of **fairness**, **consent**, and **privacy** while also guarding against potential misuse and bias in AI applications.
>
> One of the fundamental principles of ethical data usage is the respect for the privacy and autonomy of individuals whose data is collected. Collecti**ng data without express consent infringes upon an individual's right to control their personal information.** When AI datasets are compiled from data sources lacking such consent, it can lead to unintended and potentially harmful consequences. **Anonymizing data is not always sufficient, as re-identification techniques continually evolve. By ensuring that data is obtained with consent, we uphold the ethical principle of respecting individual privacy and autonomy.**
>
> Requiring express permission and fair compensation for data usage not only enhances the ethical foundations of AI but also encourages responsible development and deployment of AI technologies. **When organizations are accountable for obtaining consent and**

[260] *Comment from Chan, Maggie*, U.S. COPYRIGHT OFFICE (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-0347.
[261] *Comment from R,J*, U.S. COPYRIGHT OFFICE (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-1573 (emphasis added).
[262] *Comment from Anonymous,* U.S. COPYRIGHT OFFICE (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-3711 (last accessed Dec. 1, 2023).

compensating data creators, they are more likely to consider the ethical implications of their actions, leading to more responsible AI innovation.

Many countries have enacted data protection laws, such as the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), which mandate obtaining consent for data collection. Failing to comply with these regulations can result in legal consequences. By ensuring that AI datasets are built only with data acquired through proper consent and compensation, organizations can avoid legal challenges and maintain ethical integrity. The AI community and society at large benefit when trust and collaboration are nurtured. By respecting the rights of data creators and compensating them fairly, organizations build trust and foster goodwill. This can lead to more extensive collaboration with data creators, encouraging the sharing of high-quality datasets that advance AI research and development."[263]

C. Online News and Media Businesses are Taking Action Against OpenAI's Web Scrapers

365. Much like the average internet user, many online news and media websites are concerned that Defendants are stealing their data as well to train their AI models.

366. To combat unlicensed data collection, hundreds of publishers are trying to block AI web-crawlers from scanning their websites. Included in the list of media giants that have inserted code in an attempt to block OpenAI's web crawler, on a go forward basis, are: the New York Times, CNN, Reuters, Disney, Bloomberg, The Washington Post, ABC News, ESPN, and Insider.

367. There is increasing concern that generative AI, if it continues to grow at this rate, could greatly impact the publishing industry and even go as far as to put some newsrooms out of business. This would be ironic, given that AI's growth is and has been dependent on stealing information from these very sources.

368. News stories are a critical resource in developing generative AI. These companies' outrage demonstrates that they recognize the value of their content and believe that they should not be allowing AI web-crawlers to capitalize on that their content without paying for it in the first place. Similarly to the reactions of average internet users, these companies' response demonstrates

[263] *Comment from Anonymous*, U.S. COPYRIGHT OFFICE (Oct. 31, 2023), https://www.regulations.gov/comment/COLC-2023-0006-5788 (last accessed Dec. 1, 2023) (emphasis added).

the overarching anger towards Defendants' unfair and anticompetitive practices—spanning across the entire internet food-chain.

D. The Public is Concerned About the Legal and Long-Term Safety Implications of Normalizing Theft by Calling it "Scraping"

369. As discussed *supra*, the lethal combination of AI technology and unchecked data scraping opens the door to a wide range of dangers. Unsurprisingly, the general public has expressed fear for this technology's potentially grave capabilities.

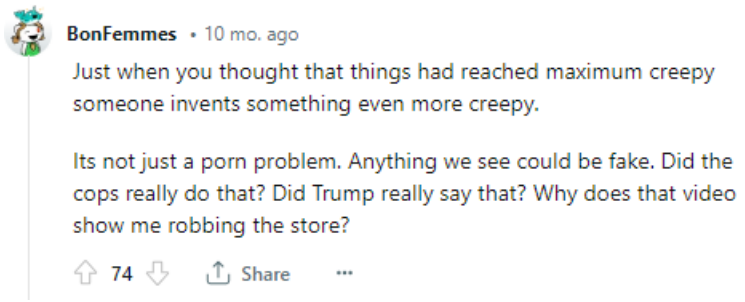
370. A X User shared her personal experience with the harms of AI and begged for change: "we need new and serious LAWS in place when it comes to AI. I've had my face put onto porn (which has caused me serious mental health issues) and now my videos are being stolen and reuploaded with others faces on it/AI."[264]

371. Recent concern has also developed around the concept of "sharenting"—parents sharing their children online.[265] Mimi Ito, a cultural anthropologist at University of California, Irvine discussed how the threat of AI makes what once was a positive experience of sharing photos of your child, negative.[266] She expressed that, "with A.I., we don't really have control of all the data that we're spewing into the social media ecosystem."[267]

[264] Tenshi (@TenshiTTV), X (Nov. 28, 2023), https://x.com/tenshittv/status/1729455572397789547?s=46&t=HHkRbC2AV14Ias3lBERw9g.
[265] Kasmir Hill, *Can You Hide a Child's Face From A.I.?*, The New York Times (Oct. 17, 2023), https://www.nytimes.com/2023/10/14/technology/artifical-intelligence-children-privacy-internet.html.
[266] *Id.*
[267] *Id.*

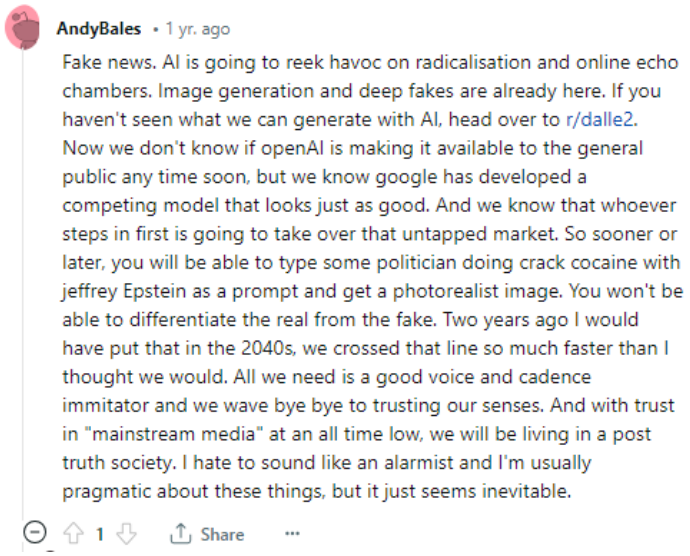
372. Others are concerned about how children can actually harm each other with this new technology. The director of the UK Safter Internet Centre addressed a recent problem schools have been having, with students using AI technology to create harmful sexual images of one another.[268] He stated: "Young people are not always aware of the seriousness of what they are doing, yet these types of harmful behaviours [*sic*] should be anticipated when new technologies, like AI generators, become more accessible to the public."[269]

373. While there are a host of concerns about how this technology could be used to harm someone's reputation, or jeopardize a child's safety—the number of internet users express a more existential concern: with AI and data scraping taking over, how are we ever supposed to know what is true and real? One Reddit user expressed this sentiment: "It[']s not just a porn problem. Anything we see could be fake. Did the cops really do that? Did Trump really say that? Why does that video show me robbing the store?"[270]



374. Another Reddit user shared that their biggest concern surrounding AI was the potential for "fake news."[271] The user elaborated on this fear: "You won't be able to differentiate the real from the fake . . . we will be living in a post truth society."[272]

[268] Tom Gerken & Joe Tidy, *Children Making AI-Generated Child Abuse Images, Says Charity*, BBC (Nov. 27, 2023), https://www.bbc.com/news/technology-67521226.
[269] *Id.*
[270] BonFemmes, *AI Deepfake Porn – We Need Legislation Passed NOW!*, REDDIT, https://www.reddit.com/r/TwoXChromosomes/comments/10q12mn/ai_deepfake_porn_we_need_legislation_passed_now/ (last visited Dec. 22, 2023).
[271] Andy Bales, *What are your Biggest Concerns About Artificial Intelligence?*, REDDIT, https://www.reddit.com/r/AskReddit/comments/vi7u4l/what_are_your_biggest_concerns_about_artificial/ (last visited Dec. 22, 2023).
[272] *Id.*


AndyBales • 1 yr. ago

Fake news. AI is going to reek havoc on radicalisation and online echo chambers. Image generation and deep fakes are already here. If you haven't seen what we can generate with AI, head over to r/dalle2. Now we don't know if openAI is making it available to the general public any time soon, but we know google has developed a competing model that looks just as good. And we know that whoever steps in first is going to take over that untapped market. So sooner or later, you will be able to type some politician doing crack cocaine with jeffrey Epstein as a prompt and get a photorealist image. You won't be able to differentiate the real from the fake. Two years ago I would have put that in the 2040s, we crossed that line so much faster than I thought we would. All we need is a good voice and cadence immitator and we wave bye bye to trusting our senses. And with trust in "mainstream media" at an all time low, we will be living in a post truth society. I hate to sound like an alarmist and I'm usually pragmatic about these things, but it just seems inevitable.

1 Share ...

375. One mother, who already was a victim of an AI scam where her daughter's voice was generated to give the impression that she was kidnapped, warned of the threat of AI altering reality.[273] She stated that if AI is "left uncontrolled, unregulated and unprotected," that it will "rewrite our understanding and perception of what is—and what is not—truth."[274]

IV. DEFENDANTS' CONDUCT VIOLATES ESTABLISHED PROPERTY AND PRIVACY RIGHTS

A. Defendants' Web-Scraping Theft

376. 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.

377. 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

[273] Yaron Steinbuch, *Traumatized Ariz. Mom Recalls Sick AI Kidnapping Scam in Gripping Testimony to Congress*, THE N. Y. POST (June 14, 2023), https://nypost.com/2023/06/14/ariz-mom-recalls-sick-ai-scam-in-gripping-testimony-to-congress/.
[274] *Id.*

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.

378. 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, forever memorized, stored, and integrated into the large language models from which the Products were born and on which they continue to run.

379. 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.[275] To create its product, Clearview scraped billions of publicly available photos from various websites and social media platforms.[276] 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 Vermont Law.[277]

380. Just like Defendants, Clearview used the stolen information to build its AI product.[278] Clearview then sold access to the product to law enforcement agencies, private companies, and other governmental agencies.[279] 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.

381. Clearview's illegal scraping practices also went undetected for years, until it was

[275] 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/.
[276] 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/.
[277] 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.
[278] *Id.*
[279] 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/.

laid bare by a New York Times expose.[280] 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.[281] Clearview was also sued by several individuals and organizations in California and elsewhere.[282]

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

383. 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. Plaintiffs 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.

1. Defendants' web scraping violates websites' terms of service that promise users data ownership and control

384. Over the course of eight (8) years, the Common Crawl dataset misappropriated by

[280] 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.
[281] *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.
[282] 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.

Defendants to train OpenAI's AI Products has scraped over 25 billion websites.[283] Among those and others Defendant scraped are countless high-traffic sites with privacy policies representing data security, terms of service promising data ownership and/or required passwords protection features.

385. Whether publicly posted or not, users maintain ownership and control of their content and data. Content creators have the right to remove their content at any time. Defendants have scraped websites, including content-centered websites, that reassure users that they maintain ownership and control of their data. For example, dropbox.com, github.com, spotify.com, reddit.com, and Microsoft.com.

386. For example, Dropbox unambiguously represents to users that, "When you use our Services, you provide us with things like your files, content, messages, contacts, and so on ("Your Stuff"). **Your Stuff is yours**."[284]

387. Github similarly assures users, ""**You retain ownership of and responsibility for Your Content**."[285]

388. Spotify's Privacy Policy also promises users "**Our legitimate interests here include protecting intellectual property and original content**."[286]

389. Reddit represents, "**You own your Contributed IP and all IP Rights in it. Nothing in the Creator Terms restricts you from exercising your IP Rights in your Contributed IP**," defining IP as "1) published and unpublished works of authorship, including audiovisual works, collective works, computer programs (including source code and object code), compilations, databases, derivative works, and literary works, 2) inventions and discoveries, improvements, machines, methods, and processes, 3) trademarks and trade names, and 4) information that is not generally known or readily ascertainable through proper means, including

[283] Elkins, *supra* note 97.
[284] *Dropbox Terms of Service*, DROPBOX (Jan. 17, 2023), https://www.dropbox.com/terms (last accessed Nov. 29, 2023).
[285] *GitHub Terms of Service*, GITHUB, https://docs.github.com/en/site-policy/github-terms/github-terms-of-service (last visited Dec. 22, 2023).
[286] *Spotify Privacy Policy*, SPOTIFY, https://www.spotify.com/ph-en/legal/privacy-policy/#8-keeping-your-personal-data-safe (last visited Dec. 22, 2023).

customer lists, ideas, and know-how."[287]

390. Accordingly, users of the websites Defendant scraped have no expectation that their content can be scraped absent their consent at any given moment.

391. Perhaps most jaw-dropping of all, is Microsoft's own empty promises to users that "Your data is **private** at work, at home, and on the go," **"You control your information,"** and "**We give you the ability to control your data, along with clear and meaningful choices over how your data is used**."[288]

392. And yet, Microsoft and OpenAI have utterly disregarded users' ownership rights to their data, using scraped content from each of these websites and more to train their AI. Defendant's conduct deprives Plaintiffs of the benefit of their contractual relationships with each of these websites—namely, it prevents these websites from being able to fulfill their promises regarding data privacy, ownership, and control.

2. Defendants' conduct violates websites' terms of service that prohibit or limit web scraping

393. In addition to interfering with the relationships of everyday internet users established by websites' terms of service, Defendants also violate their *own* contractual obligations to the websites they access—to refrain from scraping their pages.

394. Websites scraped by Defendants include provisions outright banning users from scraping the data of other users. At a minimum, most websites' terms of service drastically limit scraping—either by requiring permission or specifying that scraping not be done for a "commercial purpose." These limitations on scraping are designed to benefit the websites' entire community—including everyday internet users like Plaintiffs and the Classes, to ensure that users can share their data freely without concern for theft or misuse. The terms and conditions of a website function to regulate the actions of users, so they can maintain the safety and integrity of the entire platform

[287] *Creator Terms*, REDDIT, https://www.redditinc.com/policies/creator-terms (last visited Dec. 22, 2023).
[288] *Privacy at Microsoft*, MICROSOFT, https://privacy.microsoft.com/en-US/ (last visited Dec. 22, 2023).

for all who use it. Thousands of scraped websites prohibit web scraping, that Defendants outright ignored.

395. For example, Google, in its terms of service, references its right to suspend an account if "your ***conduct causes harm or liability to a user***, third party, or Google — for example, by hacking, phishing, harassing, spamming, misleading others, or ***scraping content that doesn't belong to you***" (emphasis added).[289]

396. Likewise, Linkedin's User Agreement requires that users "[A]gree that you will *not* . . . Develop, support or use software, devices, scripts, robots or any other means or processes (including crawlers, browser plugins and add-ons or any other technology) to ***scrape the Services or otherwise copy profiles*** and other data from the Services" (emphasis added).[290]

397. Pinterest similarly included in its terms: "In using Pinterest, ***you agree not to scrape, collect, search, copy or otherwise access data or content from Pinterest in unauthorized ways***, such as by using automated means (without our express prior permission), or access or attempt to access data you do not have permission to access" (emphasis added).[291]

398. In its terms of service, Yahoo also includes a specific prohibition on the exact type of automated scraping that Defendants engage in: "*Member conduct*. You agree not to use the Services in any manner that violates these Terms or our Community Guidelines, including to:… access or collect data, or attempt to access or collect data, from our Services using any ***automated means, devices, programs, algorithms or methodologies***, including but not limited to ***robots, spiders, scrapers, data mining tools, or data gathering or extraction tools***, for any purpose without our express, prior permission" (emphasis added).[292]

399. Because Defendants access each of website to scrape their data, Defendants are bound to the terms of service just like any other user. By web-scraping, Defendants blatantly

[289] *Terms of Service*, GOOGLE, https://policies.google.com/terms?hl=en-PH&fg=1#toc-using (last visited Dec. 22, 2023).
[290] *User Agreement*, LINKEDIN, https://www.linkedin.com/legal/user-agreement?trk=homepage-basic_footer-user-agreement (last visited Dec. 22, 2023).
[291] *Terms of Service*, PINTEREST, https://policy.pinterest.com/en/terms-of-service#section-7-termination (last visited Dec. 22, 2023).
[292] *Yahoo Terms of Service*, YAHOO, https://legal.yahoo.com/us/en/yahoo/terms/otos/index.html (last visited Dec. 22, 2023).

violate websites' provisions against this conduct. This is unlawful, unfair, and anti-competitive.

400. As a result of Defendants' misconduct, many websites have had to incorporate even more precautions to prevent Defendants from intentionally breaching terms of service and to prevent unauthorized web scraping, in order to protected users' property and privacy rights.

401. For example, in July of 2023, Twitter announced that unverified accounts will only be able to view 1,000 posts per day in order to prevent excessive data scraping.[293] Twitter went further, and as of November 2023, Twitter is not allowing individuals to view tweets unless they are logged into an account in order to make it "harder for scrapers to take Twitter's data, like ChatGPT's web browsing plugin has been doing."[294]

402. Facebook has also instituted an External Data Misuse (EDM) team of more than 100 people—including data scientists, analysts and engineers—responsible for detecting, blocking and deterring scraping. Further, Facebook employs "rate limits," designed to cap the number of times one can interact with Facebook's products during a period of time, and "data limits" to prevent people from "getting more data than they should need to use our products normally."[295]

403. TikTok's access restrictions also include rate limits and "CAPTHCAs" (designed to confirm human interaction and prevent robot access) to combat scraping.[296]

404. In addition to implementing rate limits and fake account detection defenses, LinkedIn teams "create, deploy, and maintain models and rules that detect and prevent abuse, including preventing unauthorized scraping."[297]

---

[293] Denas Grybauskas, *Will Twitter's New Rate Limits Really Stop Scraping?*, BUILTIN (Jul. 13, 2023), https://builtin.com/founders-entrepreneurship/twitter-rate-limit-scraping#.
[294] Stefanie Schappert, *Twitter Blocks Non-Users from Reading Tweets over AI Data Scraping,* CYBERNEWS (Nov. 15, 2023), https://cybernews.com/news/twitter-blocks-non-users-reading-tweets-ai-scraping/.
[295] Mike Clark, *How We Combat Scraping,* META (Apr. 15, 2021), https://about.fb.com/news/2021/04/how-we-combat-scraping/.
[296] *Why so Many Companies use TikTok Data Scrapers*, MEDIUM (Jul. 23, 2023), https://ensembledata.medium.com/why-so-many-companies-use-tiktok-data-scrapers-3b7f33c18d.
[297] Paul Rockwell, *LinkedIn Safety Series: What is Scraping?,* LINKEDIN (Jul. 15, 2021), https://blog.linkedin.com/2021/july/15/linkedin-safety-series-what-is-scraping.

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

405. 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).

406. Plaintiffs' 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, control, use, profit, sell, and exclude others from accessing or exploiting that information without consent or renumeration. *See Davis v. Facebook, Inc. (In re Facebook Inc. Internet Tracking Litig.)*, 956 F.3d 589, 598 (9th Cir. 2020) ("A right to privacy encompass[es] the individual's control of information concerning his or her person.") (internal citation omitted).

407. 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.[298]

408. A single internet user's information can be valued anywhere from $15 to $40, and even more.[299] Another study found that an individual's online identity can be sold for $1,200 on

[298] *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.
[299] *Id.*

the dark web.[300] 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.

409. 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."[301] The data forms a critical "corporate asset."

410. 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[.]"[302]

411. Because personal data is valuable personal property, market exchanges now exist where internet users like Plaintiffs and putative class members can sell or monetize their own personal data and internet usage information.[303] For example, in a study authored by Tim Morey, researchers studied the value that 180 internet users placed on keeping personal data

[300] 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.

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

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

[303] Kevin Mercandante, *10 Apps for Selling Your Data for Cash*, BEST WALLET HACKS, https://wallethacks.com/apps-for-selling-your-data/ (last updated Apr. 20, 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.

secure.[304] Contact information was valued by the study participants at approximately $4.20 per year. Demographic information was valued at approximately $3.00 per year. However, web browsing histories were valued at a much higher rate: $52.00 per year. *See true and correct summary of findings below:*

The three tiers of value

412. The value of user-correlated internet data can be quantified, because companies are willing to pay users for the exact type of information. For example, Google Inc. has a panel called "Google Screenwise Trends" which, according to the them, is designed "to learn more about how everyday people use the Internet." Upon becoming a panelist, internet users would add a browser extension that shares with Google the sites they visit and how they use them. The panelists consented to Google tracking such information for three months in exchange for one of a number of "gifts," including gift cards to retailers such as Barnes & Noble, Walmart, and Overstock.com.

413. After three months, Google also agreed to pay panelists additional gift cards "for staying with" the panel. These gift cards, mostly valued at exactly $5, demonstrated conclusively

[304] Tim Morey, *What's Your Personal Data Worth?*, DESIGN MIND (Jan. 18, 2011), https://web.archive.org/web/20131206000037/http://designmind.frogdesign.com/blog/what039s-your-personal-data-worth.html.

that internet industry participants understood the enormous value in internet users' browsing habits. Google now pays *Screenwise* panelists up to $3 per week to be tracked.[305] Similarly, another company, Facebook, has offered to *pay* users for their voice recordings.[306]

414. Now, a number of platforms have appeared where consumers can and do directly monetize their own data, and prevent tech companies, including AI companies from targeting them absent compensation and express consent. Unlike Defendants, these companies have not chosen theft to build their products, demonstrating not only harm to Plaintiffs' and the Classes' but also the unfair and illegal competitive advantage they have obtained over law-abiding competitors by not paying for or otherwise licensing content, but instead stealing it. Here are just a handful of lawful approaches by competitors, underscoring Defendants' unfair, illegal, and anticompetitive conduct:

a. **Adobe**: Adobe Firefly is Adobe's family of generative AI products.[307] Firefly is trained using Adobe Stock images—a hub that collects content that Adobe users have sold for use by Adobe and other users.[308] Adobe acknowledges the benefit that Adobe Stock content provides to its AI models, so although the Adobe Stock terms allow Adobe to freely use Adobe Stock content to train AI models, Adobe has created a Firefly bonus compensation plan to compensate Adobe Stock creators whose content was used to in AI dataset training.[309] The bonus a user earns is dependent on the number of images they submitted to Adobe Stock and the number of licenses those images accumulated.[310]

b. **Prolific**: Prolific is a platform that uses its network of participants to train AI

[305] *Cross Media Panel*, SURVEYCOOL, https://www.surveycool.com/google-cross-media-panel-review/ (last accessed Dec. 5, 2023).
[306] 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.
[307] *Firefly FAQ for Adobe Stock Contributors*, ADOBE, (Oct. 4, 2023), https://helpx.adobe.com/stock/contributor/help/firefly-faq-for-adobe-stock-contributors.html#:~:text=The%20Firefly%20bonus%20payment%20was,specific%20amount%20that%20was%20added.
[308] *Id.*
[309] *Id.*
[310] *Id.*

systems. Prolific refers to their model as "controlled data collection" because it gathers data from their "vetted collection of professional participants" who are all fairly compensated for their time and effort.[311] In turn, companies can use Prolific's data services to train their AI models, without having to engage in unethical data scraping.[312]

c. **Canva**: Canva is an online graphic design platform that allows users to create their own content. Canva has several generative AI products including Canva Assistant, Magic Media, Magic Write, and Magic Write. Canva will not use "Canva Creator" content unless they have express permission from creators—they require proactive consent from their creators to use their designs to train AI models.[313] In addition, Canva has set aside $200 million in content and AI royalties to be paid to creators who opt-in to Canva's AI training over the next three years.[314]

d. **Brave's** web browser, for example, will pay users to watch online targeted ads, while blocking out everything else.[315]

e. **The Nielsen Company**, famous for tracking the behavior of television viewers' habits, has extended their reach to computers and mobile devices through Nielsen Computer and Mobile Panel. By installing the application on your computer, phone, tablet, e-reader, or other mobile device, Nielsen tracks your activity, enters you into sweepstakes with monetary benefits, and earn points worth up to $50 per month.[316] In contrast with Defendants' theft-based AI training

---

[311] George Denison, *AI Data Scraping: Ethics and Data Quality Challenges*, PROLIFIC (Oct. 24, 2023) https://www.prolific.com/blog/ai-data-scraping-ethics-and-data-quality-challenges#:~:text=Harmful%20data%2C%20including%20abusive%20language,develop%20biases%20in%20machine%20learning ("Our platform features a minimum pay level of £6 per hour and a recommended pay level of £9 per hour").

[312] PROLIFIC, https://www.prolific.com/ai-researchers (last visited Dec. 22, 2023).

[313] *Introducing Canva Shield: Safe, Fair, and Secure AI*, CANVA, (Oct. 4, 2023) https://www.canva.com/newsroom/news/safe-ai-canva-shield/.

[314] *Id.*

[315] Brendan Hesse, *Get Paid to Watch Ads in the Brave Web Browser*, LIFEHACKER (April 26, 2019), https://lifehacker.com/get-paid-to-watch-ads-in-the-brave-web-browser-1834332279#:~:text=Brave%2C%20a%20chromium-based%20web%20browser%20that%20boasts%20an,a%20more%20thoughtful%20way%20than%20we%E2%80%99re%20accustomed%20to ("The model is entirely opt-in, meaning that ads will be disable by default. The ads you view will be converted into Brave's cryptocurrency, Basic Attention Tokens (BAT), paid out to your Brave wallet monthly").

[316] Kevin Mercandante, *Ten Apps for Selling Your Data for Cash*, BEST WALLET HACKS (June 10, 2020), https://wallethacks.com/apps-for-selling-your-data/.

model, there are currently a host of companies that offer to pay internet users to access and use their data. These companies treat data like a commodity that should be the subject of a transaction—just like any other good. Their purpose is to "benefit consumers who, until now, received nothing save targeted advertising in exchange for their data."[317]

f. **Tapestri**: Tapestri is a data collection app that allows users to generate income for sharing their data.[318] Creators of Tapestri set out to address the major issue resulting from data scraping: that consumers were being excluded from financially benefitting from the billion-dollar data industry.[319] Tapestri includes a quote from Andrew Yang, a notable technology entrepreneur, on its home page that sums up its mission: "Data is worth more than oil. And then we should be benefiting from it, not just companies."[320]

g. **ReKlaim** is a new data exchange platform that allows you to own and earn from your data.[321]

h. **BIGtoken** "is a platform to own and earn from your data. You can use the BIGtoken application to manage your digital data and identity and earn rewards when your data is purchased."[322]

415. These companies' business models *prove* that there is a legal and responsible way to collect data and train generative AI language models—one based on notice, consent, and compensation. Pay-to-use data models recognize the value of the user—for without them, there would be no data to harvest—and compensate them accordingly.

416. Indeed, Defendants' own actions prove that they are aware that there is a legal and responsible way to collect data—Defendants have initiated partnerships with Associated Press and

[317] Tatum Hunter, *These Companies will Pay you for your Data. It is a Good Deal?* THE WASH. POST (Feb. 6, 2023), https://www.washingtonpost.com/technology/2023/02/06/consumers-paid-money-data/.
[318] *About Us*, TAPESTRI, https://tapestri.io/about-us (last visited Dec. 22, 2023).
[319] *Id.*
[320] TAPESTRI, https://tapestri.io/ (last visited Dec. 22, 2023).
[321] *It's Yours*, REKLAIM, https://www.reklaimyours.com/ (last visited Dec. 22, 2023).
[322] https://bigtoken.com/faq#general_0 ("Third-party applications and sites access BIGtoken to learn more about their consumers and earn revenue from data sales made through their platforms. Our BIG promise: all data acquisition is secure and transparent, with consumers made fully aware of how their data is used and who has access to it.").

the American Journalism Project, a deal that provides these companies with attribution rights and money in exchange for their data.[323] This agreement to license data with the business entity itself shows that Defendants allow large companies to have a seat at the table, but that invitation is not extended to average internet users, whose content and personal information Defendants stole.

417. **By contrast, Defendants 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. Fair compensation should not depend on status and power. 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.

418. 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."[324]

419. 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, Plaintiffs 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.

420. In addition to monetary value, the information at issue also has non-monetary, privacy value. For example, in a recent study by the Pew Research Center, 93 percent of Americans said it was "important" for them to be "in control of who can get information" about them. Seventy-four percent said it was "very important." Eighty-seven percent of Americans said it was "important" for them not to have someone watch or listen to them without their permission.

[323] Yuvraj Malik, *Associated Press, OpenAI Partner to Explore Generative AI Use in News*, REUTERS (July 13, 2023); https://www.reuters.com/business/media-telecom/associated-press-openai-partner-explore-generative-ai-use-news-2023-07-13/.
[324] *In re Facebook, Inc. Internet Tracking Litigation*, 956 F.3d 589, 600 (9th Cir. 2020).

Sixty-seven percent said it was "very important." And 90 percent of Americans said it was "important" that they be able to "control[] what information is collected about [them]." Sixty-five percent said it was very important.[325]

421. Likewise, in a 2011 Harris Poll study, 76 percent of Americans agreed that "online companies. . . control too much of our personal information and know too much about our browsing habits."[326]

422. Consumers' sensitive and valuable personal information has increased as a commodity, where technology companies recognize the monetary value of users' sensitive, personal information, insofar as they encourage users to install applications explicitly for the purpose of selling that information to technology companies in exchange for monetary benefits.[327]

C. Defendants' Web Scraping Violated Plaintiffs' Privacy Interests

423. 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.[328]

424. 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."[329] 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

[325] *Americans' Views About Data Collection and Security*, PEW RESEARCH CENTER (May 20, 2015), https://www.pewresearch.org/internet/2015/05/20/americans-views-about-data-collection-and-security/.
[326] *Most Adults Agree Some Online Cos. Too Powerful*, MARKETING CHARTS (May 17, 2011), https://www.marketingcharts.com/industries/government-and-politics-17530/page/8?et_blog.
[327] Kari Paul, *Google Launches App that will Pay users for their Gata*, THE GUARDIAN (June 11, 2019), https://www.theguardian.com/technology/2019/jun/11/facebook-user-data-app-privacy-study; Saheli Roy Choudhury & Ryan Browne, *Facebook Pays Teens to Install an App that Could Collect all Kinds of Data*, CNBC (Jan. 30, 2019), https://www.cnbc.com/2019/01/29/facebook-paying-users-to-install-app-to-collect-data-techcrunch.html; Jay Peters, *Facebook will now Pay you for your Voice Recordings*, THE VERGE (Feb. 20, 2020), https://www.theverge.com/2020/2/20/21145584/facebook-pay-record-voice-speech-recognition-viewpoints-prounounciations-app.
[328] Geoffrey Xiao, *Bad Bots: Regulating the Scraping of Public Information*, 34(2) HARV. J.L. & TECH., 701, 706, 732 (2021).
[329] Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 493 (2006).

used in that way, to potentially reveal new, highly personal information about them.[330] If that analysis were made public, or used commercially, that would pose significant and legally cognizable privacy harms.

425. 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."[331] 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."[332]

426. 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.'"[333]

427. 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.[334]

428. Plaintiffs and the Class did not expect every iota of information they post to be scraped and fed into an AI machine learning model. To make matters worse, Defendants' ChatGPT can subsequently divulge their personal information in response to simple "attacks." As Plaintiff

[330] Xiao, *supra* note 328, at 707.
[331] Woodrow Hartzog, *The Public Information Fallacy*, 99 BOS. L. REV. 459, 515 (2019).
[332] Xiao, *supra* note 328, at 708-09.
[333] *Id*. at 709.
[334] *Id*.

Cousart explains, "this is so concerning and feels very intrusive – these are my personal details that I was sharing with friends and family… The fact that my information could be used by an external source is very concerning. I would not have posted if that was the potential future…"

429. 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.

430. Plaintiffs and the Classes did not consent to such use of their personal information. As privacy experts note**, "even if a user 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."**[335]

431. Worse, Plaintiffs 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.[336]

432. 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."[337]

433. "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

[335] *Id.* at 711.
[336] Cal. Civ. Code § 1798.99.80(d).
[337] Assemb. B. 1202, 2019-2020 Reg. Sess. (Cal. 2019) (as discussed in Xiao, *supra* note 217, at 714-715).

OpenAI did not, and still has not, registered with the State of California as required.

434. Experts acknowledge the "serious privacy harms" inherent in the type of entirely "covert information" collection in which Defendants engaged.[338] It "undermines individual autonomy and free choice."[339] 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."[340] This is not just a feeling—as described *supra*, the harm is concrete economic injury given the robust market for personal information.

435. Defendants' actions constitute a serious invasion of privacy in that they:

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

b. Violate several federal criminal laws, including the ECPA;

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

d. Invade the privacy rights of hundreds of millions of Americans (including Plaintiffs and Class Members) without their consent;

e. Constitute the unauthorized taking of valuable information from hundreds of millions of Americans; and

f. Violate Plaintiffs' and Class Members' reasonable expectation of privacy via Defendants' review, analysis, and subsequent use of Plaintiffs' and Class Members' private internet data activity that Plaintiffs and Class Members considered sensitive and confidential.

436. Committing these criminal acts against hundreds of millions of Americans—including the surreptitious and unauthorized theft of internet data of millions of Americans—constitutes an egregious breach of social norms that is highly offensive.

[338] Xiao, *supra* note 217, at 719.
[339] *Id*.
[340] *Id*.

437. Plaintiffs and Class Members now face significant distress and anxiety, stemming from the realization that Defendants have and continue to actively steal their private information, including personally identifiable information, without their informed consent or knowledge.

438. This egregious intrusion into Plaintiffs' and Class Members' private lives has not only heightened their sense of vulnerability but has also instilled a fear among the public at large. In a recent national study conducted by The Ethical Tech Project, an overwhelming majority of respondents were clearly worried about how AI products will use their data. **Results showed that 80 percent of people were concerned about AI products having access to their personal data.**[341] Additionally, Forbes cited another recent study that concluded that "**80% are concerned that their personal data is being used to train AI models**."[342] These studies underscore the harms experienced by Plaintiffs and the Classes Members here.

439. Plaintiffs' and Classes Members' awareness that their personal information, which was intended for unique audiences, is now open to unauthorized interception and analysis has disrupted their sense of security and trust in digital platforms. This distress is only exacerbated by the uncomfortable dilemma they face: either surrender their privacy to Defendants or forego the use of internet (which in today's world is impossible). Such a perpetuating cycle of unconsented use of private data has placed Plaintiffs' and Class Members in a state of perpetual vulnerability and unease, undermining their sense of security in their daily online interactions. Further, it has transformed their digital experience from a tool of empowerment into a source of anxiety and fear. This anxiety impacts Plaintiffs willingness to continue using the internet—although they want to continue sharing, posting, and accessing various websites, they only want to do so if they can ensure their data will be secure. The injunctive relief sought in this action will remedy this present harm.

---

[341] *The AI Privacy Scare: New Data Shows Americans Worry AI Products Will Abuse Their Data*, THE ETHICAL TECH Project (Oct. 24, 2023), https://news.ethicaltechproject.com/p/the-ai-privacy-scare-new-data-shows.

[342] John Koetsier, *Americans Are Terrified About AI: 80% Say AI Will Help Criminals Scam Them*, FORBES (Aug. 22, 2023), https://www.forbes.com/sites/johnkoetsier/2023/08/22/americans-are-terrified-about-data-and-ai/?sh=313853f67ca6.

440. The amount of collection of this sensitive data only exacerbates the privacy violations because when mass-harvested, the scope of the information scraped allows Defendants to assemble "digital dossiers" and comprehensive profiles of internet activity and preferences.

441. Without notice of Defendants' scraping practices, users were also denied the ability 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.[343] Instead, Plaintiffs' and the Classes' internet histories are now forever 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

442. Defendants' mass theft 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, including Plaintiffs, 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 has all been memorized and stored by Defendants' technology and thus is forever embedded into the large language models on which the Products run.

443. 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 Plaintiffs and the Classes in a zone of risk that is *incalculable* — but unacceptable by any measure of responsible data protection and use.

444. 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

[343] Xiao, *supra* note 217, at 720.

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."[344] 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."[345]

445. 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."[346]

446. Plaintiffs and the public are 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."[347] 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.[348]

447. 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.

448. 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

[344] Florian Moncomble, @coffeeseed, X (May 11, 2023), https://twitter.com/CoffeeSeed/status/1656634134616211461.
[345] Gal, *supra* note 68.
[346] Remmelt Ellen, @RemmeltE, X (Apr. 10, 2023), https://twitter.com/RemmeltE/status/1645499008075407364.
[347] Yudhanjaya Wijeratne, @yudhanjaya, X (June 9, 2023), https://twitter.com/yudhanjaya/status/1667391709679095808.
[348] Bridle, *supra* note 95.

creation of A.G.I. and then redistribute this wealth to the people."[349] 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 ongoing theft of personal information by Defendants.[350]

449. 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, Plaintiffs 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.[351]

450. 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, earlier this year, 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.

451. 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

[349] 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.
[350] *Id.*
[351] *See e.g.,* ianfinlay2000, *Time to Get Paid For Our Data?*, REDDIT, 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") (last visited Dec. 22, 2023).

as a whole is ***sprinting*** to do the same."[352] 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."[353]

452. The FTC's warning comports with FTC Commissioner Rebecca Slaughter's earlier warning, in 2021, in the Yale Journal of Law and Technology.[354] Discussing the FTC's new practice 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."[355] 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.***"[356] 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

453. 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.

454. 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.

455. *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

[352] 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).
[353] *Id.* (emphasis added).
[354] 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).
[355] *Id.*
[356] *Id.* (emphasis added).

removing their data from Defendants' for-profit commercial enterprises. Plaintiffs 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.

456. Plaintiffs 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.[357]

457. 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."[358]

458. Defendants' AI has become the virtual spy,[359] 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.

459. *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.

460. For instance, when Plaintiffs logged in to use the ChatGPT, Defendants were tracking and collecting every piece of information entered into the chatbot—including sensitive information such (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;

[357] Matt Burgess, *ChatGPT Has a Big Privacy Problem*, WIRED (Apr. 4, 2023), https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/.
[358] 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.
[359] *Id.*

(8) transaction records; (9) identifying data ChatGPT pulls from users' device or browser, like IP addresses and location; (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.

461. Plaintiffs, 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, Plaintiffs and all Class Members could not consent to Defendants' conduct because they were unaware their sensitive information would be collected and used in this manner in the first place. Thus, Defendants did not obtain *valid enforceable* consent to collect, use, and store Plaintiffs' and Class Members' sensitive information.

462. 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.

463. 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).

464. 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.[360] However, Defendants continue to train their models with the user's information – be it from the prior chats or new chats. Moreover, previously used data cannot effectively be deleted, given how the technology works, as once the language model is trained using the data, it becomes part of the model forever with no known way to "delete." 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 Defendant might decline to honor such a request, nor how long it takes to process.

465. Furthermore, as commentators have observed, Defendant OpenAI heavily pushes users not to opt out of data collection.[361] 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.

466. 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.[362]

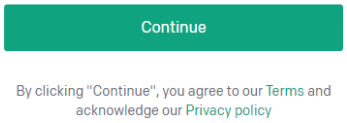
467. Defendants fail to provide accurate and comprehensive notifications to consumers about the scale of their data sharing practices. Defendants' admissions within their Privacy Policy

[360] 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 Dec. 22, 2023).
[361] 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/.
[362] 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 June 2, 2023).

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.

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



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

470. 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 Plaintiffs or the Members of

the Subclasses regarding use of these individuals' information, and no effective consent.

471. Furthermore, the language of the Privacy Policy is confusing and difficult for the average user to understand. The policy is riddled with inconsistencies and thus, as a whole, fails to paint a clear picture of what exactly users are agreeing to. Since users are unable to clearly understand Defendants' Privacy Policy and are therefore unaware of the permissions they grant Defendants by accessing the website, Plaintiffs 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

472. 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.[363]

473. In 2016, the FTC updated its publication, Protecting Personal Information: A Guide for Business, which established cybersecurity guidelines for businesses.[364] The guidelines note that 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.

474. 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

[363] *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.
[364] *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.

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.

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

476. 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.[365] 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, venders 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 by the assertion that the company both *does* and *does not* share Personal Information.

477. 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)."[366] 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.

[365] *Privacy Policy*, OPENAI, https://openai.com/policies/privacy-policy (this policy has since been updated on Nov. 14, 2023, effective Jan. 31, 2024, but this lawsuit references the version of the privacy policy that was effective at the time of the original Sept. 2023 filing).
[366] *Id.*

478. 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.

479. 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.

480. 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 under the age of 13."[367]

481. 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 Plaintiffs 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.

482. Plaintiffs and Class Members relied to their detriment on Defendants' uniform

[367] *Id.*

representations and omissions regarding data security. Now that their sensitive personal and medical information is in the possession of third parties, Plaintiffs 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.

483. Plaintiffs and Class Members trusted Defendants' Products when inputting sensitive and valuable User Data. Had Defendants disclosed to Plaintiffs 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—Plaintiffs would not have trusted Defendants' Products to input such sensitive information.

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

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

486. 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.

V. DEFENDANTS' CONDUCT POSES SPECIAL PRIVACY AND SAFETY RISKS FOR CHILDREN

487. The Products pose special risks for children, especially ChatGPT. As ChatGPT has become more pervasive and sophisticated, it has also become increasingly capable of collecting, tracking, and disclosing vast amounts of personal data about children.

488. Children's data is particularly sensitive. It can reveal not only their personal identities, but also their physical locations, habits, interests, and relationships. The indiscriminate and unauthorized collection, tracking, and disclosure of this data by powerful, profit-driven corporations undermines children's privacy and autonomy, and it also puts them at risk of abuse,

exploitation, and discrimination.

489. The safety of children in the digital environment is a foundational concern for society. According to the American Academy of Pediatrics, the overuse of social media by children places these children at heightened risk of sleep deprivation; obesity; delays in learning and social skills; decreased academic performance; behavioral problems; internet addiction; comorbid compulsive behaviors such as eating disorders; engagement in the exchange of personal, sexually explicit material; interaction with sexual predators; loss of privacy; and cyberbullying.[368]

490. Senator Michael Bennet (D-CO) recently sent a letter to the CEO of OpenAI and other industry leaders to "highlight the potential harm to younger users of rushing to integrate generative artificial intelligence (AI) in their products and services."[369] Senator Bennet wrote, "the race to deploy generative AI cannot come at the expense of our children" and that "[r]esponsible deployment requires clear policies and frameworks to promote safety, anticipate risk, and mitigate harm."[370]

491. In one illustration of the harms, Senator Bennet described how researchers prompted My AI to instruct a child how to cover up a bruise ahead of a visit from Child Protective Services.[371] Senator Bennet also detailed an incident in which My AI provided suggestions to a researcher posing as a 13-year-old girl on how to lie to her parents about an upcoming trip with a 31-year-old man.[372] It later provided suggestions for how to make losing her virginity to this man a "special experience" by "setting the mood with candles or music."[373]

492. This public introduction of AI-powered chatbot, ChatGPT, arrives during an

[368] American Academy of Pediatrics, *Beyond Screen Time: A Parent's Guide to Media Use*, PEDIATRIC PATIENT EDUC. (2021), https://doi.org/10.1542/peo_document099.
[369] Michael Bennett, *Bennett Calls on Tech Companies to Protect Kids as They Deploy AI Chatbots*, MICHAEL BENNET U.S. SEN. FOR COLO. (Mar. 21, 2023), https://www.bennet.senate.gov/public/index.cfm/2023/3/bennet-calls-on-tech-companies-to-protect-kids-as-they-deploy-ai-chatbots.
[370] *Id.*
[371] *Id*. *See also* @tristanharris, X (Mar. 10, 2023), https://twitter.com/tristanharris/status/1634299911872348160.
[372] Bennet, *supra* note 369.
[373] *Id.*

epidemic of teen mental health problems. A recent report from the Centers for Disease Control and Prevention (CDC) found that 57 percent of teenage girls felt persistently sad or hopeless in 2021, and that one in three seriously contemplated suicide.[374] In fact, the American Academy of Pediatrics (AAP), the American Academy of Child and Adolescent Psychiatry (AACAP), and the Children's Hospital Association (CHA) have declared a national emergency in child and adolescent mental health, stating that its members were "caring for young people with soaring rates of depression, anxiety, trauma, loneliness, and suicidality that will have lasting impacts on them, their families, and their communities."[375] This state of mental health across children and adults, in tandem with the increase in isolated, digital engagement is associated with normative dissociative behavior[376] and is shown to worsen depression.[377] ChatGPT exponentially exacerbates this issue by promoting human-like conversations and irresponsibly dispensing harmful, even life-threatening information—going so far as drafting suicide notes for depressed, suicidal users.[378]

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

A. Defendants Deceptively Tracked Children without Consent

494. The Children's Online Privacy Protection Act ("COPPA") requires Defendants to

[374] Moriah Balingit, *'A Cry for Help': CDC Warns of a Steep Decline in Teen Mental Health*, WASH. POST (Mar. 31, 2022), https://www.washingtonpost.com/education/2022/03/31/student-mental-health-decline-cdc/.
[375] *AAP-AACAP-CHA Declaration of a National Emergency in Child and Adolescent Mental Health*, AM. ACAD. OF PEDIATRICS (Oct. 19, 2021), https://www.aap.org/en/advocacy/child-and-adolescent-healthy-mental-development/aap-aacap-cha-declaration-of-a-national-emergency-in-child-and-adolescent-mental-health/.
[376] Amanda Baughan et al., *"I Don't Even Remember What I Read": How Design Influences Dissociation on Social Media*, 2022 CHI CONFERENCE ON HUM. FACTORS IN COMPUTING SYSTEMS (Apr. 2022), https://doi.org/10.1145/3491102.3501899.
[377] Liu Yi Lin et al., *Association Between Social Media Use and Depression Among U.S. Young Adults*, 33 DEPRESS. & ANXIETY 323, 323 (Apr. 2016).
[378] Jeremy Kaplowitz, *Man Uses ChatGPT to Write Suicide Note*, HARD DRIVE (Apr. 3, 2023), https://hard-drive.net/hd/technology/man-uses-chatgpt-to-write-suicide-note/; *see also* Gary Marcus, *The Dark Rise of Large Language Models*, WIRED (Dec. 29, 2022), https://www.wired.com/story/large-language-models-artificial-intelligence/ (GPT-3 even urged a research account to commit suicide).

obtain parental consent before monitoring, collecting, or using information from children under 13 or if they have actual knowledge that their Users are of such age. Unless Defendants obtain this consent, the law forbids collection or usage of information about these children.

495. Despite this restriction, Defendants' customary practice is to simply ignore the presence of younger Users on their application—while collecting information just like they would for an adult User—or leave it up to those Users to self-report their age, despite knowing that children can and regularly do access technology products by reporting a false birthdate.

496. Defendants are guilty of the unlawful and deceptive invasion of the right to privacy and reasonable expectation of privacy of thousands—if not millions—of children. While holding themselves out publicly as respecting privacy rights, Defendants tracked the information, behaviors, and preferences of vulnerable children solely for financial gain in violation of well-established privacy protections, societal norms, and the laws encapsulating those protections.

497. At all material times, Defendants deceived Plaintiffs and the members of the Classes and Subclasses regarding their data collection and tracking behavior. As alleged herein, Defendants knowingly and purposefully tracked, profiled, and targeted minors on the ChatGPT Platform for advertising revenue and to train LLM AI programs, like the Products. This tracking and data collection contravenes privacy rights, societal norms, and federal and state statutes, while Defendants feign compliance with these rights and statutes.

498. Defendants deceptively operated the free ChatGPT Platform as if it were only used by adults while intentionally luring thousands if not millions of children to the platform. Defendants then intentionally tracked and collected the personal information of each underage ChatGPT User (treatment to which only an adult can legally consent) in order to obtain information relevant to behavioral advertising, collect data that can be used for training the Products, and compile training datasets that can be sold to other businesses and researchers to train other AI Products. Defendants did so despite knowing that thousands if not millions of these Users were actually minor children, including children under the age of thirteen, solely for the financial benefit of Defendants, as well as their affiliates, vendors, and service providers, all of whom knowingly

and willingly consented to this unlawful conduct.

B. Defendant Designed ChatGPT to be Inappropriate for Children

499. As detailed in Section I, Defendants collect extensive data from Users to train OpenAI's language model AIs and compile training datasets.

500. Data collection of this nature requires the consent of the individual whose data is being collected. But only adults are capable of giving such consent – to the extent it was sufficiently explained in Defendant OpenAI's Privacy Policy or Terms of Service.

501. Defendant OpenAI thus inserted language into its Terms of Service and Privacy Policy which indicated that ChatGPT was intended to be used by individuals thirteen and older. More specifically, OpenAI's Privacy Policy, states the following under the heading "Children:"

> Our Service is not directed to children who are under the age of 13. OpenAI does not knowingly collect Personal Information from children under the age of 13. If you have reason to believe that a child under the age of 13 has provided Personal Information to OpenAI through the Service please email us at legal@openai.com. We will investigate any notification and if appropriate, delete the Personal Information from our systems. If you are 13 or older, but under 18, you must have consent from your parent or guardian to use our Services.[379]

502. Defendant OpenAI's Terms of Use document also references age requirements in the "Registration and Access" section, stating: "You must be at least 13 years old to use the Services. If you are under 18 you must have your parent or legal guardian's permission to use the Services."[380]

503. Defendant OpenAI prevents potential users from creating a ChatGPT user account unless the user self-reports a birthdate that indicates the user's age is thirteen or older. However, Defendants know or reasonably should know that this self-reporting of a minor child's birthdate

[379] *Privacy Policy*, OPENAI, https://openai.com/policies/privacy-policy (this policy has since been updated on Nov. 14, 2023, effective Jan. 31, 2024, but this lawsuit references the version of the privacy policy that was effective at the time of the original Sept. 2023 filing).
[380] *Terms of Use*, OPENAI, https://openai.com/policies/terms-of-use (the terms of use have since been updated on Nov. 14, 2023, effective Jan. 31, 2024, but this lawsuit references the version that was effective at the time of the original September 2023 filing).

is ineffective to keep minor children, including those under the age of 13 off the ChatGPT platform because minor users can present a false birthdate.

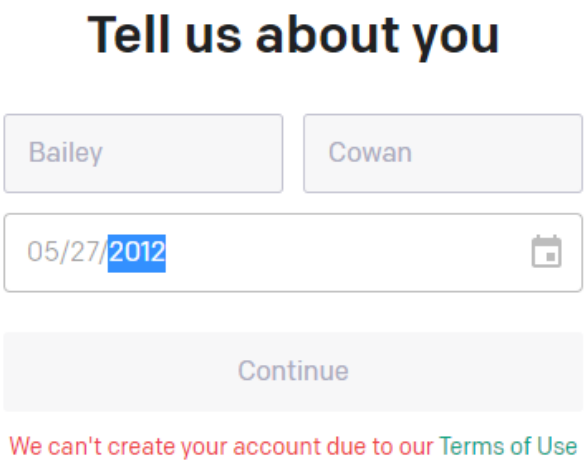
504. In fact, if a child under the age of 13 attempts to set up a ChatGPT account and is rejected for being underage, that same child can simply change their self-reported birthdate—while keeping all of their other information the same—and they will be granted immediate access. When confronted with this exact same flaw in the self-report age verification process for Snapchat, executives from Snap, Inc.—one of OpenAI's close partners[381]—admitted that such a system is effectively useless in stopping underage users from signing up for the platform.[382] Indeed, recent studies have reported that 13 percent of children ages 8–12 and 49 percent of children ages 13–17 used Snapchat in 2021 notwithstanding this same self-reporting age verification system.[383]

505. To sign up for an OpenAI account and start using ChatGPT, a child has to first provide an email and then click on a link emailed to them to verify this email. They are then directed to a page where they are asked their name and birthdate. If they enter a date of birth that indicates they are under the age of 13, they receive the following message alerting them that they cannot create an account due to OpenAI's Terms of Use.

[381] Alex Heath, *Snapchat Is Releasing Its AI Chatbot to Everyone for Free*, THE VERGE (Apr. 19, 2023) https://www.theverge.com/2023/4/19/23688913/snapchat-my-ai-chatbot-release-open-ai (Snap CEO Evan Spiegel describes the relationship between Snap and OpenAI as a "close partnership").

[382] Isobel Asher Hamilton, *Snapchat Admits Its Age Verification Safeguards Are Effectively Useless*, BUS. INSIDER (Mar. 19, 2019), https://www.businessinsider.com/snapchat-says-its-age-verification-safeguards-are-effectively-useless-2019-3#:~:text=Collins%20admitted%20that%20the%20system,mobile%20app%20is%20more%20popular.

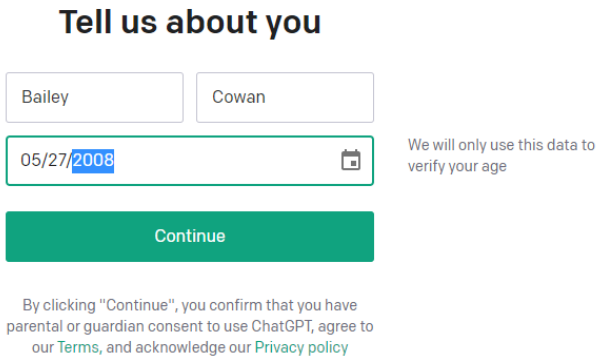
[383] Victoria Rideout et al., *The Common Sense Census: Media Use by Tweens and Teens*, COMMON SENSE MEDIA (2021), at 5, https://www.commonsensemedia.org/sites/default/files/research/report/8-18-census-integrated-report-final-web_0.pdf.



506. However, if the child then refreshes the page, they can input an earlier date of birth without any problem—and without having to use a different email or to re-verify the email address. Even though OpenAI knows that the name and email address being used has been associated with a child under the age of 13, it still allows the user to continue creating an account using this information.

507. Despite the vast amounts of data at its command, OpenAI makes no effort to verify the personal information entered, even when inconsistent information has been entered for the same user. Thus, the birthdate field is not a true age verification safeguard.

508. If the child enters a date of birth that would make them under the age of 18 but older than 13, they are able to create an account simply by hitting the Continue button.



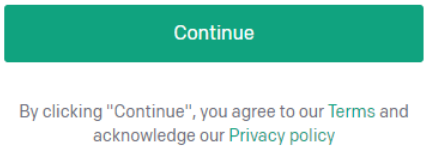
509. Under this bright green "Continue" button, there are words in small gray letters that

inform the child that, by clicking "Continue," they are confirming that they have parental or guardian consent to use ChatGPT, agree to our Terms (hyperlinked), and acknowledge our "Privacy Policy" (hyperlinked).

510. The child does not need to interact with the text in any way to indicate that they saw it or read it, much less that they read the hyperlinked Terms or Privacy Policy documents. Nor does the child user have to provide an email address for an adult so that they can confirm they actually do have parental consent and that a parent or guardian has read and agreed to the Privacy Policy and terms of use.

511. Similarly, if a user enters a date of birth that indicates they are over the age of 18, the screen looks the same, except that the small gray letters under the large green Continue button do not mention parental consent.

512. In neither case is the user asked to *agree* to the Privacy Policy – but merely acknowledge it.



C. Defendants Deprived Children of the Economic Value of their Personal Data

513. A child's personal information has equivalent (or potentially greater) value than that of an adult to companies like Defendants. First, a child is more susceptible to being influenced by advertisements as they often cannot tell the difference between content and advertisements. They also are more likely than adults to confide personal details and highly private information to ChatGPT without realizing that Defendants are using that information to train LLMs for their own financial gain, and that they may share the information with their affiliates, vendors, service providers, or partners to bolster all of these businesses' private profits.

514. Second, Defendants and/or those with whom they share User information may be

able to utilize children's personal information for the duration of their lives.[384] Plaintiffs and Minor Members of the Classes and Subclasses can no longer realize the full economic value of their personal information because it has already been collected, analyzed, acted upon, incorporated into language models, and monetized by Defendants.

515. Third, the detailed tracking of habits, preferences, thoughts, and geolocation data for young children presents unique and significant personal security and safety concerns. Quite simply, it begs the question of whether any company or its employees should have this much information about where our kids are and how to motivate their cooperation.

516. Defendants' illegal and improper collection of children's Personal Information has given them a significant "first mover" advantage that cannot be undone. ChatGPT set an unprecedented record as the fastest app to reach 100 million active users, reaching that milestone in a mere two months after its release in November 2020.

517. As a result of their unlawful conduct, ChatGPT now incorporates ill-gotten data from thousands if not millions of children who use ChatGPT without appropriate consent. The deep insights gleaned from these children's interactions with ChatGPT will enable Defendants and the for-profit companies with whom they share this data to keep children interacting with various applications, websites, language models, and platforms; to use the Personal Information of children for potentially the duration of their lives; and will solidify Defendants' dominance in the AI market by incorporating vast amounts of child-related content into Defendants' language models.

518. Publicly, Defendant OpenAI has denied marketing its ChatGPT product to children – and denied that children have utilized the application. But it is common knowledge that minors and school-aged children are using the service, as there have been widespread news reports about how schools have had to crack down on such use to prevent cheating on homework and otherwise.

[384] OpenAI's Terms of Use of ChatGPT says ChatGPT does not sell users' data to third parties. However, the terms do not disclose whether ChatGPT can display targeted advertisements to users, send third-party marketing communications, or track users based on their interactions with ChatGPT on other apps or services across the internet for advertising purposes. *See Terms of Use*, OPENAI, https://openai.com/policies/terms-of-use (these terms have since been updated on Nov. 14, 2023, effective Jan. 31, 2024, but this lawsuit references the version of the privacy policy that was effective at the time of the original Sept. 2023 filing).

Thus, Defendants knew or should have known that OpenAI's age "verification" and parental consent protocols were woefully ineffective and resulted in thousands if not millions of minor children—including those under the age of 13—gaining access to ChatGPT and sharing their personal information with the language model.

D. Defendants' Exploitation of Children Without Parental Consent Violated Reasonable Expectations of Privacy and is Highly Offensive

519. Defendant's conduct in violating privacy rights and reasonable expectations of privacy of Plaintiffs and Class and Subclass members is particularly egregious because Defendants violated social norms and laws designed to protect children, a group that is subject to such protections specifically because they are supremely vulnerable to exploitation and manipulation.

520. Parental rights to care for and control their children are fundamental liberty interests. Parental consent requirements are legally required not only to protect highly vulnerable children from deception and exploitation, but also to venerate the significant rights that parents have to determine who their children interact with and on what terms.

521. These parental rights are greatly impacted and threatened by companies like Defendants who refuse to institute reasonable and verifiable parental consent protections.

522. Though many organizations recommend against screen use for children younger than age two, 40 percent of children in this age group have used mobile media devices.[385]. Almost every family with a child younger than eight in America has a smartphone (97 percent) and/or tablet (77 percent).[386] It is exceedingly common for children to have their own devices.[387]

523. For example, a 2019 survey of media use by children aged eight through eighteen,

[385] Victoria Rideout & Michael B. Robb, *The Common Sense Census: Media Use by Kids Age Zero to Eight, 2020*, COMMON SENSE MEDIA, (2020), at 34, https://www.commonsensemedia.org/sites/default/files/research/report/2020_zero_to_eight_census_final_web.pdf.

[386] *Id*. at 15.

[387] Victoria Rideout & Michael B. Robb, *The Common Sense Census: Media Use by Tweens and Tweens, 2019*, COMMON SENSE MEDIA, (2019), at 28, https://www.commonsensemedia.org/sites/default/files/research/report/2019-census-8-to-18-full-report-updated.pdf

conducted by Common Sense Media, found that roughly 20 percent of children have a phone by the age of eight and over half (53 percent) of children in the United States have their own phone by the age of eleven.[388]

524. A survey conducted by the Center for Digital Democracy ("CDD") and Common Sense Media of over 2,000 adults found "overwhelming support for the basic principles" underpinning the Children's Online Privacy Protection Act, such as the sanctity of parental rights to care for and control their children from online surveillance and from commercial co-optation of their children's data.[389] Of the parents polled: 75 percent strongly disagreed with the statement "[i]t is okay for advertisers to track and keep a record of a child's behavior online if they give the child free content;" 84 percent strongly disagreed with the statement that "[i]t is okay for advertisers to collect information about a child's location from that child's mobile phone;" and 89 percent strongly agreed with the statement "[b]efore advertisers put tracking software on a child's computer, advertisers should receive the parent's permission."[390] 93 percent of the parents surveyed indicated that a federal law requiring online sites and companies to ask parents' permission before they collect Personal Information from children under age 13 was "a good idea."[391] Against this backdrop, Defendants' knowing exploitation of children without adequate parental involvement is not only illegal but also highly offensive to social norms and mores.

CLASS ALLEGATIONS

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

a. **Non-User Class**: All persons in the United States whose PII, Personal Information, or Private Information was disclosed to, or accessed, collected,

[388] *Id.* at 4.
[389] Center for Digital Democracy & Common Sense Media, *Survey on Children and Online Privacy: Summary of Methods and Findings*, https://democraticmedia.org/assets/resources/COPPA-Executive-Summary-and-Findings-1635879421.pdf (last visited Dec. 22, 2023).
[390] *Id.*
[391] *Id.*

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.

f. **Minor ChatGPT User Class.** All persons in the United States who, while 16 years or younger, used ChatGPT, or other platforms, programs, or applications which integrated ChatGPT API or ChatGPT Plug-In, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

g. **Minor Microsoft User Class:** All persons in the United States who, while 16 years or younger, 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.

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 User SubClass:** All persons within the State of California who used ChatGPT, or other platforms, programs, or applications, which integrated ChatGPT API or ChatGPT Plug-In, 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 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.

iv. **California Microsoft User.** All persons within the State of California 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.

v. **California Minor User SubClass**: All persons within the State of California who, while 16 years or younger, used ChatGPT, or other platforms, programs, or applications which integrated ChatGPT API or ChatGPT Plug-In, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

The Illinois Subclass

i. **Illinois NonUser Subclass:** All persons within the State of Illinois whose PII, Personal Information, or Private Information was disclosed to, or accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

i. **Illinois User Subclass:** All persons within the State of Illinois who used ChatGPT, Microsoft programs, or other platforms, programs, or applications, which integrated ChatGPT API or ChatGPT Plug-In, whose Private Information was disclosed to, or intercepted, accessed, collected, tracked, taken, or used by Defendants without consent or authorization.

ii. **Illinois Biometric Subclass:** All persons within the State of Illinois whose biometric information was disclosed to, or accessed, collected, taken, or used by Defendants without consent or authorization.

526. **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) Plaintiffs' counsel and Defendants' counsel; and (6) the legal representatives, successors, and assigns of any such excluded persons.

527. Plaintiffs reserve 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. Plaintiffs reserve the right under Federal Rule of Civil Procedure 23(c)(4) to seek certification of particular issues.

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

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

530. **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.

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

532. **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 are 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. California Constitution right to privacy;

2. Comprehensive Computer Data Access and Fraud Act;

3. Federal and California Wiretapping Act (ECPA and CIPA);
4. California Unfair Competition Law, Bus. & Prof Code § 17200;
5. New York General Business Law §§ 349, *et seq.*
6. Illinois Biometric Information Privacy Act § 740 ILCS 14;
7. Illinois Consumer Fraud & Deceptive Trade Practices Act;

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

1. Common law Negligence;
2. A violation of the Constitutional Right to Privacy under California law;
3. Conversion;

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, 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 Open AI 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, 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. Comprehensive Computer Data Access and Fraud Act, Cal. Penal Code § 502, *et seq.*
3. California Invasion of Privacy Act;
4. California Constitution Right to Privacy;
5. California Unfair Competition Law, Bus. & Prof Code §§ 17200;
6. Common Law Negligence;
7. Conversion.

f) Whether as a result of Defendants' collection, scraping, and uses of the protected Private Information, ChatGPT User, ChatGPT Plug-In User, 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 result of Defendants' interception, collection, receipt, or unauthorized uses of Private Information, ChatGPT User, ChatGPT Plug-In User, 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.

533. **Typicality:** Plaintiffs' claims are typical of the claims of other Class Members in that Plaintiffs 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.

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

535. The declaratory and injunctive relief sought in this case includes, by way of example and without limitation:

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 Plaintiffs and the Class 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.

536. **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 Plaintiffs 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' unauthorized disclosure of Users' sensitive information was negligent;
b) Whether Defendants owed a duty to Plaintiffs' and the Class not to disclose their sensitive user information to unauthorized third parties;
c) Whether Defendants breached their duty to Plaintiffs' and the Class not to disclose their sensitive user information to unauthorized third parties;
d) Whether Defendants represented to Plaintiffs and the Class that they would protect Plaintiffs' and the Members of Classes and Subclasses Private Information;
e) Whether Defendants violated Plaintiffs' and Classes' right to privacy;
f) Whether Defendants violated federal and/or California wiretapping laws (ECPA/CIPA);
g) Whether Plaintiffs and the Class 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; and

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

537. **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.

538. 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 PLAINTIFFS' AND CLASS MEMBERS' NON-STATUTORY CLAIMS

539. Courts "have permitted the application of California law where the plaintiffs' 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).

540. 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.

541. 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).

542. Defendant Open AI is headquartered in California; this is where Defendant Open AI's nerve center of its business operations is located. This is where Defendant Open AI 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.

543. Furthermore, Defendant Open AI requires that California law applies to disputes between Defendant Open AI and ChatGPT Users.

544. 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 Plaintiffs and the Classes than any other state, and the state most intimately concerned with the claims and outcome of this litigation.

545. 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 Open AI is headquartered in California, and Defendant Microsoft conducts business (at least as it relates to Defendant Open AI) in California, there is no conflict in applying California

law to non-resident consumer claims.

546. 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.

547. 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, Microsoft User Classes against Defendants)

548. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

549. The Federal Wiretap Act, as amended by the Electronic Communications Privacy Act of 1986 (the "ECPA"), 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.

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

a. The computer codes and programs that Defendants use to track the Plaintiffs' and the Classes' communications;

b. The Plaintiffs' and the Classes' browsers and applications;

c. The Plaintiffs' and the Classes' computing and mobile devices;

d. Defendants' web servers;

e. The web servers of websites from which Defendants tracked and intercepted the Plaintiffs' and the Classes' communications;

f. The computer codes and programs used by Defendants to effectuate their tracking and interception of the Plaintiffs' and the Classes' communications;

g. The plan that Defendants carried out to effectuate its tracking and

interception of the Plaintiffs' and the Classes' communications.

551. The ECPA protects both the sending and reception of communications.

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

553. 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.

554. Defendants' actions were done contemporaneously with the Plaintiffs' and the Classes' sending and receiving those communications.

555. Defendants' interception included "contents" of electronic communications made from Plaintiffs and the Class 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 Plaintiffs and the Class sent to those websites and for which Plaintiffs received communications in return from those websites.

556. The transmission of data between Plaintiffs and the Class on the one hand and the websites and other web properties other than Defendants' on which Defendants tracked and intercepted Plaintiffs' and the Classes' 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 ECPA. 18 U.S.C. § 2510(12).

557. 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 ECPA. 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.

558. None of the Defendants were authorized parties to the communications because Plaintiffs and the Class were unaware of the collection and interception. Neither can Defendants

manufacture their own status as parties to the communications by surreptitiously intercepting those communications.

559. Both 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.

560. **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**: Plaintiffs' 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**: Plaintiffs' 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**: Plaintiffs' 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 Open AI;

561. **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).

562. Plaintiffs, 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.

563. **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).

564. Defendants intentionally accessed, and obtained access to the contents of Plaintiffs', 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 Plaintiffs, the Classes, and Subclasses to access the protected computers.

565. **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) Open AI website and mobile application(s) for ChatGPT.

566. **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) Plaintiffs' and Classes', Subclasses' computing devices (Mac and Windows

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

(2) Plaintiffs' and Classes', 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, Programs which have integrated ChatGPT API. [Microsoft User Class is Excluded]

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

568. The transmissions of Plaintiffs', 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).

569. 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.

570. As Plaintiffs 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 Plaintiffs and ChatGPT API Class Members, and/or third-party entities.

571. In disregard for Plaintiffs', and ChatGPT API Class Members' privacy rights,

Defendants act as a third-party "eavesdropper," redirecting Plaintiffs and ChatGPT API Class Members' electronic communications to Defendants' own servers for appropriation, and training of their Products.

572. Defendants' interception of the contents of Plaintiffs' and ChatGPT API Class Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Plaintiffs' 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 Plaintiffs and ChatGPT API Class Members also to be able to understand, collect, and use for training Plaintiffs' and ChatGPT API Class Members' communications.

573. **Unauthorized Purpose**. Plaintiffs and ChatGPT API Class Members did not authorize Defendants to acquire, access, or intercept the content of their communications on third party platforms, websites, applications. Therefore, such interception and recording of communications invades Plaintiffs', and ChatGPT API Class Members' privacy. Defendants intentionally intercepted the contents of Plaintiffs' 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.

574. **While in Transmission.** Through this calculated scheme of using ChatGPT API to intercept, acquire, transmit, and record Plaintiffs' 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 Plaintiffs, and ChatGPT API Class. 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.

575. In sending and in acquiring the content of Plaintiffs', 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 Plaintiffs, 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).

576. Plaintiffs, individually, on behalf of the GPT API Class and Subclass Members, seek 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

577. The allegations for violation of 18 U.S.C. § 2510 arising out of Defendant Microsoft's interception of Plaintiffs, ChatGPT User Class Members' communications which occurred on ChatGPT platform.

578. The transmissions of Plaintiffs', 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).

579. 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.

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

581. In disregard for Plaintiffs' and ChatGPT User Class Members' privacy rights, Defendant Microsoft acts as a third-party "eavesdropper," redirecting Plaintiffs' and ChatGPT

User Class Members' electronic communications to Defendant Microsoft's own servers for appropriation, and training of their Products.

582. Defendant Microsoft's interception of the contents of Plaintiffs', ChatGPT User Class Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Defendant Open AI or third-party entities. As described above, the 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 Plaintiffs, and ChatGPT User Class Members, but also to be able to understand, collect, and use for training Plaintiffs' and ChatGPT User Class Members' communications.

583. **Unauthorized Purpose**. Plaintiffs 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, applications. Moreover, Plaintiffs 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 Plaintiffs', ChatGPT User Class Members' privacy. Defendant Open AI illegally divulged the content of such communications to Defendant Microsoft. Defendant Microsoft intentionally intercepted the contents of Plaintiffs' 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.

584. **While in Transmission.** Through this calculated scheme of using ChatGPT technology to intercept, acquire, transmit, and record Plaintiffs', 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 Plaintiffs 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.

585. In sending and in acquiring the content of Plaintiffs', 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 Plaintiffs, 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).

586. Plaintiffs, individually, on behalf of the ChatGPT User Class Members, seek 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 Open AI's Interception of Microsoft User Class Members which occurred on Microsoft's Websites, Platforms, Applications, Programs which have integrated ChatGPT.

587. The allegations for violation of 18 U.S.C. § 2510 arising out of Defendant Open AI'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.

588. The transmissions of Plaintiffs' 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).

589. 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.

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

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

592. Defendant Open AI interception of the contents of Plaintiffs', Microsoft Subclasses Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Plaintiffs', 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 Plaintiffs, Microsoft Subclasses Members, but also to be able to understand, collect, and use for training Plaintiffs', Microsoft Subclasses Members' communications.

593. **Unauthorized Purpose**. Plaintiffs and Microsoft Subclasses did not authorize Defendant Open AI 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 Plaintiffs', Microsoft Subclasses Members' privacy. Defendant Open AI intentionally intercepted (and continues to intercept) the contents of Plaintiffs', 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.

594. **While in Transmission.** Through this calculated scheme of using ChatGPT API to

intercept, acquire, transmit, and record Plaintiffs', Microsoft Subclasses Members' electronic communications, Defendant Open AI 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 Plaintiffs' 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.

595. In sending and in acquiring the content of Plaintiffs', and Class Members' communications with third-party platforms, Defendant Open AI's purpose was tortious, and designed to violate federal and state laws. By intentionally using, or endeavoring to use, the contents of Plaintiffs' 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 Open AI violated and continues to violate 18 U.S.C. § 2511(1)(a).

596. Plaintiffs, individually, on behalf of the Microsoft Subclasses Members, seek 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: VIOLATIONS OF THE COMPREHENSIVE COMPUTER DATA ACCESS AND FRAUD ACT ("CDAFA"), CAL. PENAL CODE § 502, *et seq.*

(on behalf of Non-User Class and California Non-User SubClass)

597. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

598. Cal. Penal Code § 502 provides: "For purposes of bringing a civil or a criminal action under this section, a person who causes, by any means, the access of a computer, computer system, or computer network in one jurisdiction from another jurisdiction is deemed to have personally accessed the computer, computer system, or computer network in each jurisdiction."

599. Smart phone devices with the capability of using web browsers and applications are "computers" within the meaning of the statute.

600. Tablet devices with the capability of using web browsers and applications are

"computers" within the meaning of the statute.

601. Laptop and desktop computing devices with the capability of using web browsers and applications are "computers" within the meaning of the statute.

602. Each Plaintiff is the owner of Private Information, and his/her data at issue.

603. Defendants violated Cal. Penal Code § 502(c)(2) by knowingly accessing and without permission taking, copying, analyzing, and using Plaintiffs' and Class Members' Private Information.

604. Each Plaintiff, as a direct and proximate result of Defendants' unauthorized access and taking, copying, analyzing, and using Plaintiffs' and Class Members' Private Information, each Plaintiff and Class Member was harmed.

605. Defendants were unjustly enriched, by acquiring their sensitive and valuable Private Information without permission and using it for their own financial benefit to advance its AI development business. Plaintiffs and Class Members retain a stake in the profits Defendants earned from their Private Information and other internet contributions (*i.e.*, data) because, under the circumstances, it is unjust for Defendants to retain those profits.

606. Defendants accessed, scraped, copied, analyzed, and used Plaintiffs' and Class Members' Private Information and other internet contributions (*i.e.*, data) without authorized consent, in and from the State of California, where Defendants: (1) maintain at least one principal place of business wherein the activities were contemplated, planned, and executed therefrom; (2) accessed, scraped, copied, analyzed, and used the Plaintiffs' and Class Members' data at issue; (3) used servers that provided access to the scraped webpages from which Defendants accessed and scraped Plaintiffs' and Class Members' data. Accordingly, Defendants caused the access of Plaintiffs' and Class Members' data from California and are therefore deemed to have accessed Plaintiffs' and Class Members' data in California. *See* Cal. Pen. Code § 502(c)(2) (**an entity can violate the CDAFA by "knowingly access[ing] and without permission tak[ing], cop[ying], or mak[ing] use of any data."**) (emphasis added).

607. As a direct and proximate result of Defendants' unlawful conduct within the

meaning of Cal. Penal Code § 502, Defendants have caused harm to Plaintiffs and Class Members and have been unjustly enriched in an amount to be proven at trial.

608. Plaintiffs, on behalf of themselves and Class Members, seek compensatory damages and/or disgorgement in an amount to be proven at trial, and declarative, injunctive, or other equitable relief.

609. Plaintiffs and Class members are entitled to punitive or exemplary damages pursuant to Cal. Penal Code § 502(e)(4) because Defendant's violations were willful and, upon information and belief, Defendants are guilty of oppression, fraud, or malice as defined in Cal. Civil Code § 3294.

610. Plaintiffs and the Class Members are also entitled to recover their reasonable attorneys' fees pursuant to Cal. Penal Code § 502(e).

COUNT THREE: VIOLATION OF THE CALIFORNIA INVASION OF PRIVACY ACT ("CIPA"), CAL. PENAL CODE § 631, *et seq.*

(on behalf of All Plaintiffs and the ChatGPT Class, ChatGPT API User Classes, Microsoft User Classes, and California User Subclasses against Defendants)

597. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

598. 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.

599. 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 . . . .

600. 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 . . . .

601. Under either section of the CIPA, a defendant must show it had the consent of all parties to a communication.

602. 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.

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

604. All Plaintiffs used ChatGPT.

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

606. The following items constitute "machine[s], instrument[s], or contrivance[s]"

under the CIPA, and even if they do not, Google's 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 the Plaintiffs' and Class members' communications;

b. The Plaintiffs' and the Classes' browsers and mobile applications;

c. The Plaintiffs' and the Classes' computing and mobile devices;

d. Defendants' web and ad servers;

e. The web and ad-servers of websites from which Defendants tracked and intercepted the Plaintiffs' and the Classes' communications;

f. The computer codes and programs that Defendants used to effectuate tracking and interception of the Plaintiffs' and the Classes' communications; and

g. The plan Defendants carried out to effectuate the tracking and interception of the Plaintiffs' and the Classes' communications.

607. The data collected by Defendants constituted "confidential communications," as that term is used in Section 632, because Plaintiffs and the Class had objectively reasonable expectations of privacy that the information would not be used for Defendants' AI products.

608. Plaintiffs and the Class 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.

609. Pursuant to California Penal Code § 637.2, Plaintiffs and the Class 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.

610. Plaintiffs bring this claim individually and on behalf of the members of the proposed Class against Defendants.

611. 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).

612. 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).

613. 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 which occurred on Third-Party Websites, Platforms, Applications, Programs which have integrated ChatGPT API. [Microsoft User Subclass is Excluded]

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

615. The transmissions of Plaintiffs' 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, websites which integrate ChatGPT API (i.e., Stripe, Snapchat, etc.) qualify as "electronic communications" under Cal. Penal Code §629.51(2).

616. 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.

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

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

619. Defendants' interception of the contents of Plaintiffs' and ChatGPT API Class Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Plaintiffs' 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 Plaintiffs and ChatGPT API Class Members', but also to be able to understand, collect, and use for training Plaintiffs' and ChatGPT API Class Members' communications.

620. 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 Plaintiffs' 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 Plaintiffs 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.

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

622. Plaintiffs, individually, on behalf of the GPT API Class, also seek 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

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

624. The transmissions of Plaintiffs' 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).

625. 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.

626. As Plaintiffs and ChatGPT User Class Members ask questions, or otherwise interact with Defendant Open AI, Defendant Open AI intentionally aids and abets Defendant Microsoft to intentionally tap and intercept, electrically or otherwise, the lines of internet communications of Plaintiffs' and Chat-GPT User Class Members' searches and communications.

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

628. Defendant Microsoft's interception of the contents of Plaintiffs' and ChatGPT User Class Members' communications happens contemporaneously with their exchange of such communications, whether such communications are directed to Defendant Open AI 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 Plaintiffs and Chat-GPT User Members, but also to be able to understand, collect, and use for training Plaintiffs' and Chat-GPT User Class Members' communications.

629. 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.

630. Through this calculated scheme of using ChatGPT to intercept, acquire, transmit, and record Plaintiffs' 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 Plaintiffs and Chat-GPT User Class. 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.

631. In sending and in acquiring the content of Plaintiffs' 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 Plaintiffs, 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).

632. 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.")

633. Plaintiffs, individually, on behalf of the GPT ChatGPT User Class Members, seek 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 Open AI's Interception of Microsoft User Class Members which occurred on Microsoft's Websites, Platforms, Applications, Programs which have integrated ChatGPT.

634. The allegations for violation of CIPA § 631(a) arise out of Defendant Open AI'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.

635. The transmissions of Plaintiffs' 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, websites which integrate ChatGPT API qualify as "electronic communications" under Cal. Penal Code §629.51(2).

636. 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.

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

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

639. Defendant Open AI's interception of the contents of Plaintiffs' and Microsoft Subclasses Members' communications happens contemporaneously with their exchange of such communications on Microsoft platforms, whether such communications are directed to Plaintiffs'

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

640. 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.")

641. Plaintiffs, individually, on behalf of the Microsoft Subclasses Members, seek 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.

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

643. Plaintiffs 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 All Plaintiffs and the Classes against Defendants)

644. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

645. As discussed above, Plaintiffs believe that California law should apply to all claimants, including out of state residents.

646. 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

647. 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 the Plaintiffs' and Nationwide 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 used by Plaintiffs and the Class, including but not limited to Reddit, Twitter, TikTok, Spotify, YouTube, Facebook, WhatsApp, and other websites, without their consent. Defendants' illegal web scraping violates privacy laws, California civil and criminal cyberstalking laws, and other laws outlined in this complaint.

b) Defendants failed to register as data brokers under California law as required: As

discussed *supra*, in allegations 301-305, Defendants violated California law requiring that those who acquire personal information through scraping practices register as data brokers. As defined by California law, a "data broker" is a business that collects and sells personal data of consumers with whom the business does not have a "direct relationship" with. Cal. Civ. Code § 1798.99.80. Any business that mees the definition of a "data broker" is required to register with the Attorney General. *Id.* at § 1798.99.82. OpenAI qualifies as a "data broker," because the company scrapes the internet to collect personal information of consumers who it does not otherwise have a business relationship with, and then uses that data to train its commercial AI products, such as ChatGPT. Despite their data brokering practices, OpenAI has failed to register as such with the California Attorney General.

c) 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.

d) Open AI's Interception of Communications and Accessed, Collected, and Tracked Private Information on Microsoft Platforms: Defendant Microsoft aided Defendant Open AI 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 Open AI 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.

e) 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.

f) Defendants Interference with Plaintiffs' Contractual Relationships with Websites: Through its web-scraping conduct, Defendants unlawfully interfered with Plaintiffs contractual relationships with the websites they accessed and shared personal data with. Defendants web-scraping prevented the websites from upholding their contractual obligations to Plaintiff, since these websites' terms of service and privacy policies promised that Plaintiffs would maintain control and ownership of their data.

g) Defendants Breached their Own Contractual Obligations with the Websites they Scraped: Since Defendants accessed and interacted with the websites they scraped, they, like any other internet user, were subject to a contractual relationship with the websites they scraped. Defendants scraping practice violated the terms of service and privacy policies of websites explicitly banning or limiting web-scraping. Because these anti-scraping policies are designed to benefit the entire platform's community, and protect the safety and data of all users, Defendants conduct harmed Plaintiffs, who were intended third-party beneficiaries of these contracts.

648. 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.

649. Defendants' conduct violates the EPCA, CFAA, CDAFA, 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.,* CalOPPA, Cal. Pen. Code § 484 and §532, California Bus. & Prof. Code § 22576, and other tort claims stated in this lawsuit. The violations of EPCA, CFAA, CDAFA, CIPA, and other tort claims stated in this lawsuit, are incorporated herein by reference.

650. 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).

651. "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).

652. As alleged, Defendant uses web scraping technology to collect information from webpages across the internet and, in so doing, Defendant gathers and compiles personal information about consumers that is reflected on those webpages.

653. 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, when 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 Plaintiffs and affected Nationwide Classes 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).

654. Defendant's failure to provide notice to Plaintiffs and Nationwide Classes 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).

655. 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).

656. 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).

657. 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).

658. 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).

659. 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.[392] Open AI'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."[393]

660. 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.[394]

661. 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 . . ."[395] Users are instructed that, "to the extent applicable under local law, [they] can exercise privacy rights. . . by submitting a request to dsar@openai.com."[396]

662. 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 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.

663. CalOPPA applies to Defendant OpenAI because it operates a commercial website and online service that collects personally identifiable information about individual consumers residing in California. Cal. Bus. & Prof. Code § 22575(a).

664. CalOPPA defines personally identifiable information as first and last name; home or other physical address, including street name and name of a city or town; e-mail address; telephone number; social security number; any other identifier that permits the physical or online contacting of a specific individual; information concerning a user that the website or online service

[392] *Privacy Policy*, OPENAI, https://openai.com/policies/privacy-policy (this policy has since been updated on Nov. 14, 2023, effective Jan. 31, 2024, but this lawsuit references the version of the privacy policy that was effective at the time of the original Sept. 2023 filing).
[393] *Id.*
[394] *Id.*
[395] *Id.*
[396] *Id.*

collects online from the user and maintains in personally identifiable form in combination with an identifier described in this subdivision. Cal. Bus. & Prof. Code § 22577(a).

665. OpenAI violates CalOPPA because while its Privacy Policy instructs consumers regarding how they can review and request changes to OpenAI's collection of their data, the disclosures in this regard are misleading and incomplete in that they do not disclose that data used to train the Products realistically cannot be deleted from the Products.

666. OpenAI also violates CalOPPA by failing to disclose whether other parties may collect personally identifiable information about an individual consumer's online activities over time and across different Web sites when a consumer uses the OpenAI's website of ChatGPT service.

667. Furthermore, OpenAI also violates CalOPPA by knowingly collecting information from minors under the age of thirteen ("13") without appropriate measures to ensure parental consent and without ensuring that the full deletion of information about minors is feasible from their products.

668. By failing to fulfill their contractual obligations under their Privacy Policy (which was expressly incorporated in the Terms of Use, Defendants also failed to confer on Plaintiffs the benefit of the bargain, thereby causing them economic injury. This breach is a violation of California Business and Professions Code § 22576, which prohibits a commercial website operator from "knowingly and willfully" or "negligently and materially" failing to comply with the provisions of its posted Privacy Policy. *See* Cal. Bus. and Prof. Code § 22576. Defendants' also violated terms of the various websites by scraping Plaintiffs' data, where scraping is expressly prohibited within the terms and conditions of such websites. (See Exh. 1).

669. Plaintiffs, individually and on behalf of the Nationwide Classes seek: (i) an injunction requiring OpenAI to revise its Privacy Policy to include reasonable protections for children and Minors User Subclass, to fully disclose all information required under CalOPPA and COPPA, and to delete all information previously collected in violation of these laws; (ii) 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; (iii) relief under Cal. Bus. & Prof. Code § 17200, *et seq*., including, but not limited to, restitution to Plaintiffs and other members of the Nationwide 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, (iv) reasonable attorney's fees (pursuant to Cal. Code of Civ. P. § 1021.5).

670. 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.

671. As a direct and proximate result of Defendants' misconduct, Plaintiffs and Nationwide Classes 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.

672. As a result of Defendants' unlawful conduct, Plaintiffs and Nationwide Classes 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

673. 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.

674. 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 continued to be used to train Defendants' Products, without consent of the users.

675. 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.

676. Defendants' conduct is unfair under each of these tests. As described above, Defendants' conduct violates the policies underlying privacy laws and, with respect to children under the age of thirteen, the mandates of COPPA and CalOPPA. The gravity of the harm of Defendants' illegal scraping, interception and misuse of Private information to train their AI Products, as well as secret tracking, profiling, and targeting of children is significant and there is no corresponding benefit to consumers of such conduct.

677. Finally, because Plaintiff N.B. and Minor User Subclass Members were minors unable to consent to or understand Defendants' conduct—and because their parents did not consent to this conduct and were misled by their belief that Defendants would follow applicable laws and societal expectations about children's privacy as well as Defendants' statements—they could not have avoided the harm.

678. 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.

679. Defendants have intentionally and deceptively misled parents and the public about

Defendants' intention to use the ChatGPT language model and its free chatbox application to attract children in order to gain access to the Personal Information of such children and to exploit such children's Personal Information for Defendants' financial gain.

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

681. Defendants' representations and omissions were material because they were likely to deceive reasonable consumers such as the parents or guardians of Plaintiff N.B. and Class and Subclass Members about the terms under which their children were interacting with the ChatGPT app as well as the fact that Defendant was collecting and profiting from minors' Personal Information without their parents and guardians' knowledge or consent.

682. Defendants had a duty to disclose the above-described facts due to the important public interest in securing the privacy of minors' Personal Information and the fact that minors are unable to fully protect their own interests.

683. 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."

684. The expectations of Plaintiff N.B.'s parents and guardians included that Defendants would not track their children's online activity, without their consent, in order for Defendants to reap huge profits from building out the fastest growing application ever, and the most advanced AI language models of all time.

685. The parents and guardians of Plaintiff N.B. and Minor User Subclass members reasonably expected that Defendants respected children's privacy online, in accordance with societal expectations and public policy as well as state and federal statutes and regulations including COPPA, CalOPPA, and Federal Trade Commission regulations.

686. At the same time, Defendants have, at all times throughout the Class Period, been well aware that children, including children under the age of 16 and under the age of 13, access ChatGPT; have actively sought to increase engagement with ChatGPT by children; and have

sought to exploit, for commercial purposes and gain, thousands if not millions of minor users of ChatGPT.

687. Defendants' knowledge of the widespread use of ChatGPT by children and failure to disclose that they are tracking, profiling, and targeting such children and/or profiting from this behavior, while at the same time representing that OpenAI and ChatGPT comply with law and societal expectation, and does not permit and does not seek to reach children, are likely to and, in fact, did deceive Plaintiff N.B. and Minor User Subclass Members and their parents or guardians. Defendants' conduct therefore constitutes deceptive business practices in violation of Cal. Bus. & Prof. Code §17200.

688. Additionally, to the extent that Defendants have represented to Plaintiff N.B., Minor User Subclass members, and their respective parents and guardians that Defendants can and will disclose to such individuals, upon request, the private information that Defendants have gathered about any such minor user or non-user, and that such information can be deleted, these representations are fraudulent and deceptive because it is functionally impossible for Defendants to "undo" the fact that their LLMs have learned on this private information and incorporated that learning in such a manner that the information cannot be meaningfully segregated, identified, extracted, and deleted.

689. 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 Plaintiffs' 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 N.B.'s and Class Members' User Data without their authorization or knowledge in order to profit off of the information, and to deliver advertisements to Plaintiffs and Class Members, among other unlawful purposes.

690. 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.

691. Defendants have unlawfully tracked, targeted, and profiled minor Plaintiffs, and Minor User Subclass Members without obtaining parental consent in violation of COPPA, CalOPPA, Federal Trade Commission regulations, and other laws.

692. Defendants also engaged in business acts and practices deemed "unlawful" under the UCL as to the Nationwide Classes by unlawfully tracking, targeting, and profiling Plaintiffs' minor children, in violation of the California Constitution.

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

694. Further, Defendants' business model was inconsistent with common practice. As discussed *supra*, there are several other data collection and AI training companies that acquire data in ethical and legal ways. These company's practices—including paying consumers in exchange for voluntarily sharing their data—prove that Defendants practices are unlawful and unfair toward competition. Were Defendants to have implemented these lawful business practices, Plaintiffs and Class Members not only would have had a choice over whether to share their data, but they would have economically benefitted from doing so.

695. 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.

696. As a direct and proximate result of Defendants' misconduct, Plaintiffs and Nationwide Classes 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.

697. As a result of Defendants' unlawful conduct, Plaintiffs and Nationwide Classes 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. Defendants conduct causes ongoing injury to Plaintiffs and the Nationwide Classes Members—namely, Defendants harmful web-scraping has, and continues to have, a chilling effect on Plaintiffs' and Nationwide Classes Members' continued use of the internet.

698. Plaintiff N.B. and Minor User Subclass Members placed trust in Defendants as major and reputable companies that represented they were in compliance with applicable laws and societal interests in safeguarding minors' Personal Information.

699. Additionally, Defendants had the sole ability to understand the extent of their collection of Personal Information, and the parents or guardians of Plaintiff N.B. and Minor User Subclass Members could not reasonably have discovered—and were unaware of—Defendants' secret tracking, profiling, and targeting.

700. Defendants invaded Plaintiff N.B.'s and Minor User Subclass Members' privacy without their or their parents and guardians' consent.

701. Because Defendants held themselves out as complying with law and public policy regarding minors' privacy rights, the parents or guardians of Plaintiff N.B. and California Minor User Subclass Members acted reasonably in relying on Defendants' misrepresentations and omissions.

702. Plaintiff N.B. and Minor User Subclass Members could not have reasonably avoided injury because Defendants' business acts and practices unreasonably created or took advantage of an obstacle to the free exercise of their decision-making. By withholding the important information that it was collecting and profiting from minors' Personal Information, Defendants created an asymmetry of information.

703. Further, Defendants' conduct is immoral, unethical, oppressive, unscrupulous and substantially injurious to Plaintiffs and Nationwide Classes Members and Minor User Subclass Members, and there are no greater countervailing benefits to consumers or competition.

704. Plaintiffs, as well as the Nationwide Classes Members and Minor Subclass 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 Plaintiffs and Nationwide Classes Members and Minor Subclass Members.

705. As a direct and proximate result of Defendants' deceptive acts and practices, Plaintiffs and Nationwide Classes Members and Minor Subclass 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 Plaintiffs and Nationwide Classes Members and Minor Subclass Members.

706. Plaintiffs' and Nationwide Classes Members' and Minors Members' Personal Information has tangible value; it is now in the possession of Defendants, who has used and will continue to use it for financial gain.

707. Plaintiffs' and Nationwide Classes Members and Minor Subclass Members injury was the direct and proximate result of Defendants' conduct described herein.

708. Defendants' retention of Plaintiffs' and Nationwide Classes Members' Personal Information presents a continuing risk to them as well as the general public.

709. Plaintiffs, individually and on behalf of the Nationwide Classes Members and Minor Subclass Members, seek: (1) an injunction requiring Defendants to permanently delete, destroy or otherwise sequester the Private Information collected without consent; (2) compensatory restitution of Plaintiffs' and Nationwide Classes 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).

710. Had Plaintiffs and Nationwide Classes Members and Minor Subclass Members known Defendants would disclose and misuse their User Data in contravention of Defendants' representations, they would not have used Defendants' Products.

711. 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.

712. As a direct and proximate result of Defendants' misconduct, Plaintiffs and Nationwide Classes Members and Minor Subclass 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.

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

COUNT FIVE

VIOLATION OF ILLINOIS'S BIOMETRIC INFORMATION PRIVACY ACT, 740 ILCS 14/1, *et seq.*

(on behalf of Plaintiff Roberts and Illinois Biometric Subclass against Defendants)

714. Plaintiff Roberts, individually and on behalf of the Illinois Biometric Subclass, repeats and re-alleges the allegations contained in the paragraphs 1 through 18, 52 through 60, and 146 through 538 as if fully set forth herein.

715. BIPA created statutory duties for Defendants with respect to the collection of biometric identifiers and biometric information of the Plaintiff Roberts and the Illinois Biometric Subclass members.

716. Defendants violated BIPA section 15(b)(1) by systematically collecting the Plaintiff Roberts' and the Illinois Biometric Subclass' biometric identifiers and biometric identifiers, by taking their photos off of the internet and scanning their facial geometry and related biometric information to train the algorithms on which DALL-E runs, without first informing the Plaintiff Roberts and the Illinois Biometric Subclass.

717. In so doing, Defendants also violated section 15(b)(2) of BIPA by not informing Plaintiff Roberts and the Illinois Biometric Subclass members in writing of the purpose for their collection of facial geometry and related biometric information, and by failing to inform them in

writing of the length of time Defendants would collect their biometric identifiers and biometric information, including scans of their facial geometry and related biometric information.

718. Defendants violated section 15(b)(3) of BIPA by not receiving a written release executed by Plaintiff Roberts and the Illinois Biometric Subclass, the subjects of the biometric identifiers and biometric information.

719. Section 15(c) of BIPA makes it unlawful for any private entity to among other things, "sell, lease, trade, or otherwise profit from a person's or a customer's biometric identifier or biometric information," which Defendants did by incorporating that information into the Products for commercial gain. Without the facial scans and recording of facial geometry of Plaintiff Roberts and the Illinois Biometric Subclass, DALL-E could not exist.

720. BIPA prohibits private entities "in possession of a biometric identifier or biometric information" from "disclos[ing], redisclos[ing], or otherwise disseminat[ing] a person's or a customer's biometric identifier or biometric information unless" any one of four enumerated conditions are met. 740 ILCS 14/15(d)(1)-(4). None of such conditions are met here.

721. Defendants disclose, redisclose and disseminate, and at all relevant times disclosed, redisclosed and disseminated, the Plaintiff Roberts 's and the Illinois Biometric Subclass' "biometric identifiers," including but not limited to their face geometry scans, and "biometric information" without the consent of any of them or their "legally authorized representatives." 740 ILCS 14/15(d)(1). Moreover, the disclosures and redisclosures did not "complete[] a financial transaction requested or authorized by" the Plaintiff Roberts , the Illinois Biometric Subclass or any of their legally authorized representatives. 740 ILCS 14/15(d)(2). Nor are, or at any relevant times were, the disclosures and redisclosures "required by State or federal law or municipal ordinance." 740 ILCS 14/15(d)(3). Finally, at no point in time were the disclosures ever "required pursuant to a valid warrant or subpoena issued by a court of competent jurisdiction." 740 ILCS 14/15(d)(4).

722. BIPA mandates that a private entity "in possession of biometric identifiers or biometric information" "develop a written policy, made available to the public, establishing a

retention schedule and guidelines for permanently destroying biometric identifiers and biometric information when the initial purpose for collecting or obtaining such identifiers or information has been satisfied or within 3 years of the individual's last interaction with the private entity, whichever occurs first." 740 ILCS 14/15(a).

723. But Defendants do not publicly provide any written policy establishing any retention schedule or guidelines for permanently destroying the Plaintiff Roberts 's and the Illinois Biometric Subclass members' "biometric identifiers" and "biometric information." 740 ILCS 14/15(a).

724. BIPA also commands private entities "in possession of a biometric identifier or biometric information" to: (1) store, transmit, and protect from disclosure all biometric identifiers and biometric information using the reasonable standard of care within the private entity's industry; and (2) store, transmit, and protect from disclosure all biometric identifiers and biometric information in a manner that is the same as or more protective than the manner in which the private entity stores, transmits and protects other confidential and sensitive information. 740 ILCS 14/15(e). Based on the facts alleged herein, including Defendants' lack of a public written policy, their failure to inform Users that Defendants obtain such users' "biometric identifiers" and "biometric information," their failure to obtain written consent to collect or otherwise obtain Users' "biometric identifiers" and "biometric information," and their unauthorized dissemination of Users' "biometric identifiers" and "biometric information," Defendants have also violated this provision.

725. Plaintiff Roberts and the Illinois Biometric Subclass have been directly harmed by these violations. They have been deprived of their control over valuable information, and otherwise suffered monetary and non-monetary losses. By depriving them of control over their valuable information, Defendants misappropriated the value of their biometric identifiers and biometric information, and are profiting from this unlawful conduct.

726. Plaintiff Roberts and the Illinois Biometric Subclass seek (i) injunctive and equitable relief requiring Defendants to comply with BIPA; (ii) statutory damages of $5,000 per

intentional or reckless violation of BIPA and statutory damages of $1,000 per negligent violation of BIPA; and (iii) reasonable attorneys' fees and costs and other litigation expenses as permitted by statute. 740 ILCS 14/20(1)-(4).

COUNT SIX

ILLINOIS CONSUMER FRAUD AND DECEPTIVE BUSINESS PRACTICES ACT

815 ILL. COMP STAT. §§ 505, *et seq.*

(on behalf of Plaintiff Roberts and Illinois Subclasses against Defendants)

450. Plaintiff Roberts, individually and on behalf of the Illinois Subclasses, repeats and re-alleges the allegations contained in the paragraphs 1 through 18, 52 through 60, and 146 through 538 as if fully set forth herein.

451. Defendant OpenAI and Defendant Microsoft are "persons" as defined by 815 Ill. Comp. Stat. §§ 505/1(c).

452. The Plaintiff Roberts and Illinois Subclasses Members are "consumers" as defined by 815 Ill. Comp. Stat. §§ 505/1(e).

453. Defendants' conduct as described herein was in the conduct of "trade" or "commerce" as defined by 815 Ill. Comp. Stat. § 505/1(f).

454. Defendants' deceptive, unfair, and unlawful trade acts or practices, in violation of 815 Ill. Comp. Stat. § 505/2, include:

a) Defendants have exploited Non-Users and Users of their Products, by stealing such individuals' data at scale from web crawler caches without permission from the data owners and without any way of segregating out any given Non-Users' or User' data from the datasets used to train OpenAI's LLMs upon request of such individuals—including where such individuals are minors.

b) Defendants knew that they were collecting and/or profiting from individuals' Personal Information and that the risk of collecting of such Personal Information was highly likely. Defendants' actions in engaging in the above-named deceptive acts and practices were negligent, knowing and willful, and/or wanton and reckless with respect

to the rights of the Plaintiff Roberts and members of the Illinois Subclasses;

c) As described herein, Defendants are misrepresenting that they have and are complying with common law and statutory duties pertaining to the security and privacy of the Plaintiff Roberts 's and Illinois Subclass Members' data, including but not limited to duties imposed by the FTC Act, 15 U.S.C. § 45, the Illinois Uniform Deceptive Trade Practices Act, 815 Ill. Comp. Stat. § 510/2(a), and the Illinois Personal Information Act, 815 Ill. Comp. Stat. §§ 530/10(a).

d) As described herein, Defendants have and are omitting, suppressing, and concealing the material fact that they are stealing and profiting from the mass collection and analysis of the Plaintiff Roberts 's and Illinois Subclasses Members' data at scale and without adequate or effective consent; and

e) Omitting, suppressing, and concealing the material fact that they did not comply with common law and statutory duties pertaining to the security and privacy of Plaintiff Roberts 's and the Illinois Subclass Members' data, including but not limited to the fact that they are functionally unable to delete such data once it has been incorporated into their LLMs as training data.

455. Defendants' representations and omissions were material because they were likely to deceive reasonable consumers about the terms of use of the Products, as well as the available mechanisms for seeking to exert control over Plaintiff Roberts 's and Illinois Subclasses Members' data.

456. Defendants intended to mislead the Plaintiff Roberts and Illinois Subclasses Members and induce them to rely on their misrepresentations and omissions.

457. The above unfair and deceptive practices and acts by Defendants were immoral, unethical, oppressive, and unscrupulous. These acts caused substantial injury that these consumers could not reasonably avoid; this substantial injury outweighed any benefits to consumers or to competition.

458. Defendants acted intentionally, knowingly, and maliciously to violate Illinois's

Consumer Fraud Act, and recklessly disregarded Plaintiff Roberts 's and Illinois Subclasses Members' rights.

459. As a direct and proximate result of Defendants' unfair, unlawful, and deceptive acts and practices, the Plaintiff Roberts and Illinois Subclasses Members have suffered and will continue to suffer injury, ascertainable losses of money or property, and monetary and non-monetary damages, as described herein.

460. Plaintiff Roberts and Illinois Subclasses Members seek all monetary and non-monetary relief allowed by law, including damages, restitution, punitive damages, injunctive relief, and reasonable attorneys' fees and costs.

COUNT SEVEN

ILLINOIS CONSUMER FRAUD AND DECEPTIVE BUSINESS PRACTICES ACT 815 ILL. COMP. STAT. §§ 510/2, *et seq.*

(on behalf of Plaintiff Roberts and Illinois Subclasses against Defendants)

461. Plaintiff Roberts, individually and on behalf of the Illinois Subclasses, repeats and re-alleges the allegations contained in the paragraphs 1 through 18, 52 through 60, and 146 through 538 as if fully set forth herein.

462. Defendant OpenAI and Defendant Microsoft are "persons" as defined by 815 Ill. Comp. Stat. §§ 510/1(5).

463. Defendants engaged in deceptive trade practices in the conduct of its business, in violation of 815 Ill. Comp. Stat. §§ 510/2(a), including:

a) Representing that goods or services have characteristics that they do not have, 815 Ill. Comp. Stat. § 510/2(a)(5);

b) Representing that goods or services are of a particular standard, quality, or grade if they are of another, 815 Ill. Comp. Stat. § 510/2(a)(7);

c) Advertising goods or services with intent not to sell them as advertised, 815 Ill. Comp. Stat. § 510/2(a)(9); and

d) Engaging in other conduct that creates a likelihood of confusion or misunderstanding,

815 Ill. Comp. Stat. § 510/2(a)(12).

464. Defendants' deceptive acts and practices include those enumerated, *supra*, in paragraph 454.

465. Defendants' representations and omissions were material because they were likely to deceive reasonable consumers about the terms of use of the Products, as well as the available mechanisms for seeking to exert control over the Plaintiff Roberts 's and Illinois Subclasses Members' data.

466. Defendants intended to mislead the Plaintiff Roberts and Illinois Subclasses Members and induce them to rely on its misrepresentations and omissions.

467. The above unfair and deceptive practices and acts by Defendants were immoral, unethical, oppressive, and unscrupulous. These acts caused substantial injury that these consumers could not reasonably avoid; this substantial injury outweighed any benefits to consumers or to competition.

468. Defendants acted intentionally, knowingly, and maliciously to violate Illinois's Consumer Fraud Act, and recklessly disregarded Plaintiff Roberts's and Illinois Subclasses Members' rights.

469. As a direct and proximate result of Defendants' unfair, unlawful, and deceptive acts and practices, the Plaintiff Roberts and the Illinois Subclasses Members have suffered and will continue to suffer injury, ascertainable losses of money or property, and monetary and non-monetary damages, as described herein.

470. The Plaintiff Roberts and Illinois Subclasses Members seek all monetary and non-monetary relief allowed by law, including damages, restitution, punitive damages, injunctive relief, and reasonable attorneys' fees and costs.

COUNT EIGHT: NEGLIGENCE

(on behalf of All Plaintiffs and the Classes against Defendants)

727. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

728. Defendants owed a duty to Plaintiffs 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.

729. 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 Plaintiffs 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 Plaintiffs and the Classes in a zone of risk that is incalculable – but unacceptable by any measure of responsible data protection and use.

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

731. Plaintiffs and Classes' 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.

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

733. The damages suffered by Plaintiffs and the Classes' 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.

471. Defendants' negligence directly caused significant harm to Plaintiffs and the Classes.

COUNT NINE: INVASION OF PRIVACY

(on behalf of All Plaintiffs and the Classes against Defendants)

734. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

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

736. Plaintiffs' and the Classes' 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).

737. The phrase "and privacy" was added in 1972 after voters approved a proposed legislative constitutional amendment designated as Proposition 11. In a 1975 California Supreme Court opinion interpreting Proposition 11, the Court concluded that the legislative intent behind adopting this amendment was to curb businesses' control over the unauthorized collection and use of consumers' personal information.[397] The Court highlighted a state election brochure circulated by proponents of the amendment in coming to this conclusion:[398]

> 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 the information gathered for one purpose in order to serve 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. (Emphasis in original).

738. The principal purpose of this constitutional right was to protect against unnecessary

[397] *White v. Davis*, 13 Cal. 3d 757, 774 (1975) ("the moving force behind the new constitutional provision was a more focussed [sic] privacy concern, relating to the accelerating encroachment on personal freedom and security caused by increased surveillance and data collection activity in contemporary society").
[398] *Id*.

information gathering, use, and dissemination by public and private entities, including Defendants.

739. To plead a California constitutional invasion of 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.

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

a. Scraped data from password protected websites or websites designated for a specific audience;[399]

b. Scraped websites' Terms and Conditions as alleged herein;

c. Intercepted data pursuant to the ECPA and California Wiretap Act as alleged herein;

d. 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*;

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

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

741. Plaintiffs and the Class had a reasonable expectation of privacy under the circumstances in that Plaintiffs and the Class could not reasonably expect Defendants would commit acts in violation of federal and state civil and criminal laws, such as theft of their data.

742. Defendant's actions constituted a serious invasion of privacy in that it:

a. Invaded a zone of privacy protected by the Fourth Amendment, namely the right to privacy in data contained on personal computing devices, including

[399] *See U.S. Department of Justice v. Reporters Committee for Freedom of the Press*, 489 U.S. 749, 763-64, 109 S. Ct. 1468, 103 L. Ed. 2d 774 (1989) ("information may be classified as 'private' if it is 'intended for or restricted to the use of a particular person *or group* or *class of persons*: not freely available to the public.'") (emphasis added) (quoting Webster's Third New International Dictionary 1804 (1976)).

web search and browsing histories;

b. Violated several federal criminal laws, including the ECPA;

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 Plaintiffs and the Class) without their consent;

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

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

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

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

745. Defendants' intentional intrusion into Plaintiffs' and the Classes' 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.

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

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

748. 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 Plaintiffs' and the Classes' reasonable expectations of privacy.

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

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

751. Plaintiffs and the Class have been damaged by Defendants' invasion of their privacy and are entitled to injunctive relief.

COUNT TEN: CONVERSION

(on behalf of All Plaintiffs and the Classes against Defendants)

752. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

753. Property is the right of any person to possess, use, enjoy, or dispose of a thing, including intangible things such as data or communications. Plaintiffs' and Nationwide Classes Members' personal information is their property. *Calhoun v. Google* LLC, 526 F. Supp. 3d 605, 636 (N.D. Cal. 2021).

754. Defendants unlawfully collected, used, and exercised dominion and control over the Nationwide Classes Members' personal and private information without authorization.

755. Defendants wrongfully exercised control over Plaintiffs' and Nationwide Classes' information and have not returned it.

756. Plaintiffs and Nationwide Classes Members have been damaged as a result of Defendants' unlawful conversion of their property.

COUNT ELEVEN: UNJUST ENRICHMENT

(on behalf of All Plaintiffs and the Classes against Defendants)

757. Plaintiffs hereby incorporate paragraphs 1 through 547 as if fully stated herein.

758. 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 Plaintiffs and Nationwide Classes Members for the commercial training of its ChatGPT and other AI language models.

759. 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 Plaintiffs and the members of the Nationwide Classes.

760. As a result of Defendants' conduct, Plaintiffs and Nationwide Classes Members suffered actual damages in the loss of value of their Private Information and the lost profits from the use of their Private Information.

761. It would be inequitable and unjust to permit Defendants to retain the enormous economic benefits (financial and otherwise) they have obtained from and/or at the expense of Plaintiffs and Classes Members.

762. Defendants will be unjustly enriched if they are permitted to retain the economic benefits conferred upon them by Plaintiffs and Nationwide Classes 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 Plaintiffs and Nationwide Classes members.

763. Plaintiffs and Nationwide Classes members are therefore entitled to recover the amounts realized by Defendants at the expense of Plaintiffs and Nationwide Classes Members.

764. Plaintiffs and the Nationwide Classes have no adequate remedy at law.

765. Plaintiffs and the members of the Nationwide 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, Plaintiffs on behalf of themselves and the Proposed Classes that they seek to represent, 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 Plaintiffs to represent the Classes;

C. Appoint undersigned counsel to represent the Classes;

D. Award compensatory damages (including treble damages, where appropriate) to

Plaintiffs and the Class 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 Plaintiffs and the Class against Defendants;

F. Award nominal damages to Plaintiffs and the Class 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 Plaintiffs and the Class 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;

J. Award Plaintiffs and the Class their reasonable costs and expenses incurred in this Action, including attorneys' fees, costs, and expenses; and

K. Grant Plaintiffs and the Class such further relief as the Court deems appropriate.

JURY TRIAL DEMANDED

Plaintiffs demand a jury trial on all triable issues.

DATED: December 29, 2023

/s/ Michael F. Ram

MORGAN & MORGAN
COMPLEX LITIGATION GROUP
Michael F. Ram
John A. Yanchunis
Ryan J. McGee