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
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 Plaintiff and the Proposed Class

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

PLAINTIFF JILL LEOVY, individually, and on behalf of all others similarly situated,

Plaintiff,

vs.

GOOGLE LLC,

Defendant.

Case No. 3:23-cv-3440-AMO

SECOND AMENDED CLASS ACTION COMPLAINT

1. DIRECT COPYRIGHT INFRINGEMENT

DEMAND FOR JURY TRIAL

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

TABLE OF CONTENTS

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

Plaintiff Jill Leovy ("**Plaintiff**"), individually and on behalf of all others similarly situated, brings this action against Defendant Google, LLC ("**Defendant**" or "**Google**"). Plaintiff's allegations are based upon personal knowledge as to herself and her own acts, and upon information and belief as to all other matters.

INTRODUCTION

1. The Constitution and the Copyright Act recognize the critical importance of protecting the authors and creators' exclusive rights over their works. The legal protection of copyrighted materials is intended to nourish and encourage innovation and creativity. As the United States Supreme Court has declared: "The immediate effect of our copyright law is to secure a fair return for an 'author's creative labor. But the ultimate aim is, by this incentive to stimulate artistic creativity for the general public good."[1] Knowing that their works are protected, creators are more likely to invest time and effort in creating new works, leading to a richer and more diverse literary landscape.

2. Google, however, has elected to disregard the Constitution and the Copyright Act, and grant itself a license to steal copyright protected works created by millions of Americans to build, train, and commercialize its Artificial Intelligence ("AI") Products like "Gemini" (previously known as "Bard"), the chatbot Google released to compete with OpenAI's "ChatGPT." For years, Google secretly harvested a massive quantity of pirated and copyrighted works, including a trove of books, articles, images, photographs, and millions of other protected works. Google used the protected works to create its AI Products, and by doing so, increased its market share by billions of dollars, unjustly profiting off the infringed intellectual property. Specifically, Google posted that it achieved revenue of $80.5 billion dollars in the first quarter of 2024—which reflects a 15% increase from the first quarter of 2023, when it initially released Gemini (originally introduced as Bard).[2]

[1] *Sony Corp. of Am. v. Universal Studios, Inc.*, 464 U.S. 417, 432, 78 L. Ed. 2d 574, 104 S. Ct. 774 (1984).

[2] Nick robins-Early, *Alphabet Hails Once in a Lifetime AI Opportunity as Revenue Rises*, THE GUARDIAN (April 25, 2024), https://www.theguardian.com/technology/2024/apr/25/google-revenue-quarter-one.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

3. Generative AI Products like Gemini are designed to understand and generate human language. They are characterized by their size and complexity, and are able to write human-like responses, articles, books, and other works, mimicking the expressive works on which they were built.

4. Creative and expressive works are critical to the AI training process because this is how products like Gemini learn to "create" works. This mass theft of personal and copyrighted information has stunned internet users around the world.

5. The FTC issued a stern warning to the AI industry in May 2023 regarding this sudden sprint to collect as much training data as they can find: "Machine learning is no excuse to break the law . . . The data you use to improve your algorithms must be lawfully collected . . . companies would do well to heed this lesson."[3]

6. Rather than heed the FTC's warning and stop its years-long theft of data, Google elected to copy and download the works of writers and other creators, without compensation, to build and train its AI Products like Gemini, capable of performing now the same type of work for which these writers and authors would be paid.

7. Defendant's billion-dollar AI Products' success was possible only because it copied, downloaded, and digested the protected copyrightable works contained in billions of actual texts across the internet – all without paying a dime to creators.

8. Defendant's excuse to the massive infringement is that it believes that everything on the internet is a fair game for Google to take for private gain and commercial use, as it announced to the public through the updated version of its online privacy policy from July 2023. However, that's not the law. The law requires Google to obtain the creators' consent to reproduce (by copying, downloading, and storing the works accessed from the online databases or websites) and pay fair value for such use.

[3] *Statement of Commissioner Alvaro M. Bedoya Joined by Chair Lina M. Khan and Commissioner Rebecca Kelly Slaughter*, FEDERAL TRADE COMMISSION (MAY 31, 2023), available at https://www.ftc.gov/legal-library/browse/cases-proceedings/public-statements/statement-commissioner-alvaro-m-bedoya-joined-chair-lina-m-khan-commissioner-rebecca-kelly-slaughter-0.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

9. Authors and other creators are outraged to learn that Google has been using their works to train its AI Products. This reaction is understandable, given Google's history of encroaching on intellectual and other data property rights. Its claims of ownership over all internet content, including copyrighted works, strikes many as bold and insidious—because it is.

10. Google has since invited the world to engage in "dialogue" about what data collection and protection efforts should look like in the new era of AI, while continuing to steal and infringe the works of authors to improve and expand its AI Products. That invited a backlash of its own, naturally, as a classic case of too little too late. One commentator aptly translated Google's "invitation" into the truth: "Now that we've already trained our LLMs on all your proprietary and copyrighted content, we will finally start thinking about giving you a way to opt out of any of your future content for being used to make us rich."[4]

11. Google had options other than to steal copyrighted information. Google could have paid to license the copyrighted material it stole, so as not to infringe on the owner's exclusive rights. The legal acquisition of such material depends on consent and consideration.

12. There are companies that specialize in curating and selling datasets for AI training purposes that contain information obtained with the *express consent* of the content creators or subjects of the personal or copyrighted information. Using these datasets might be more expensive than stealing, but using this data has one critical advantage: it is legal. Against this backdrop, Google's decision to instead take copyrighted material without notice, consent, or fair compensation not only violates the individual rights of millions, but also gives Google an unfair advantage over smaller competitors who lawfully license copyrighted material.

13. As part of its theft, Google illegally accessed restricted, subscription-based websites to take the content of millions without permission and infringed on millions of materials explicitly protected by copyright, including previously stolen property from websites known for pirated collections of books and other creative works. Without this mass theft of copyrighted information belonging to real people, many of Google's AI products, including Gemini, would not exist.

[4] Matt G. Southern, *Google Calls For Public Discussion On AI Use Of Web Content*, SEARCH ENGINE JOURNAL (July 7, 2023), https://www.searchenginejournal.com/google-calls-for-public-discussion-on-ai-use-of-web-content/491053/ (quoting Barry Adams, Twitter/X, since deleted).

Defendant, on a mass scale, unlawfully reproduced valuable intellectual property of creators without paying a penny for the use of these creative works to build its AI Products. Defendant must be enjoined from these ongoing violations of copyright laws. It must also be ordered to either delete the data already obtained illegally or obtain licenses to use the copyrighted material it has stolen.

PARTIES

Plaintiff Jill Leovy ("Plaintiff Leovy")

14. Plaintiff Leovy is a New York Times best-selling author and investigative journalist residing in the State of Texas.

15. Plaintiff Leovy owns the registered copyright in this book, which includes customary copyright-management information including the name of the author and the year of publication (2015). The registered copyright owned by Plaintiff Leovy is included as **Exhibit A.** Defendant misappropriated Plaintiff Leovy's award-winning non-fiction book called *Ghettoside: A True Story of Murder in America*, by illegally copying the book in full without her knowledge or consent to train "Gemini" and Google's other AI Products.

16. The copyrighted work that Defendant misappropriated and otherwise infringed reflects over a decade of Plaintiff Leovy's investigative journalism and work, including novel insights on a topic few have researched and written about in as much detail. As a result of Defendant's large-scale theft of copyrighted materials, Plaintiff Leovy was never paid a penny for Google's unauthorized reproduction of the book, on which it developed its multi-billion-dollar product.

17. Defendant's infringement thus deprives creators like Leovy of deserved royalties, and undermines their financial stability and incentive to create new works. Absent the relief sought in this Action, Plaintiff Leovy and millions of authors and creators like her presently have no ability to demand Google to stop copying/reproducing their works, and/or provide fair compensation for the use of their works.

Defendant

18. **Defendant Google LLC** is headquartered in Mountain View, California. It was founded in 1998 as an American search engine company. Google's search business now amounts

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

to $149 billion, with over 85 percent market share in the global desktop search engine market worldwide. In 2015, as part of its corporate restructuring, Google LLC became a subsidiary of its newly formed parent company, Alphabet, Inc. Google LLC is currently one of the world's largest for-profit tech companies, specializing in internet related services and products with a special emphasis on "web-based search and display advertising tools, search engine, cloud computing, software, and hardware."[5]

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

JURISDICTION AND VENUE

20. This Court has subject matter jurisdiction under 28 U.S.C. § 1331 because this case arises under the Copyright Act, 17 U.S.C. § 501.

21. 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 Google LLC is headquartered in this District, Defendant gains significant revenue and profits from doing business in this District, Class Members affected by Defendant's copyright infringement reside in this District, and Defendant employs numerous people in this District—a number of whom work specifically on making decisions regarding the handling and use of copyrighted materials. 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. Defendant's 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.

22. The Court has general personal jurisdiction over Defendant, because Defendant is

[5] *Google LLC*, BLOOMBERG, https://www.bloomberg.com/profile/company/8888000D:US#xj4y7vzkg (last visited June 27, 2024).

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

headquartered in California and makes decisions concerning the Product(s) and the use of use of copyrighted materials from California. Defendant also advertises and solicits business in California.

FACTUAL BACKGROUND

I. GOOGLE'S DEVELOPMENT OF GOOGLE GEMINI.

23. Defendant's AI Product, Google Gemini, operates as an advanced language model, capable swaths of information in response to users' questions and prompts.[6] Its user interface is presented as "a dialogue box where users type in their queries."[7] When users enter their questions in to Gemini's dialogue box, Gemini uses the treasure trove of information it has been trained on and information available on the internet to provide users with responses—often admittedly plagiarizing the writing of others.[8] Gemini is able to respond to users not only with text-based answers, but also via image-based answers, adding another function to its capabilities.[9]

24. Gemini was initially built on the LaMDA LLM.[10] Google has since transitioned Gemini to PaLM 2,[11] a LLM trained on 3.6 trillion tokens (strings of words), more powerful than any existing model.[12] Due to its vast training data, Gemini can also replicate and mimic the human writing – reproducing the works of the artists, authors, and creators on whose content it was trained on. The end result is that Gemini is not only built on the works of millions of creators and authors,

---

[6] Andy Patrizio, *Google Bard*, TECHTARGET, https://www.techtarget.com/searchenterpriseai/definition/Google-Bard (last visited June 27, 2024).

[7] Ben Wodecki, *Google Unveils Bard: Its Version of ChatGPT*, AI BUS. (Feb. 7, 2023), https://aibusiness.com/google/google-unveils-bard-its-version-of-chatgpt.

[8] *See* Avram Piltch, *Google Bard Plagiarized Our Article, Then Apologized When Caught*, TOM'S HARDWARE (March 23, 2023), https://www.tomshardware.com/news/google-bard-plagiarizing-article. The author of this article questioned Google Gemini, at the time Google Bard, about computer processors, and Gemini provided an answer that was word for word taken form a Tom's Hardware article. *Id.* Then, when asked if Gemini had just plagiarized the article, Gemini responded, "yes what I did was a form of plagiarism." *Id.*

[9] Sabrina Ortiz, *What is Google Bard? Here's Everything You Need to Know*, ZDNET (June 1, 2023), https://www.zdnet.com/article/what-is-google-bard-heres-everything-you-need-to-know/.

[10] Joe Jacob, *What Sites Were Used for Training Google Bard AI?*, MEDIUM (Feb. 11, 2023), https://medium.com/@taureanjoe/what-sites-were-used-for-training-google-bard-ai-1216600f452d.

[11] Ortiz, *supra* note 9.

[12] Jennifer Elias, *Google's Newest A.I. Model Uses Nearly Five Times More Text Data for Training than Its Predecessor*, CNBC (May 17, 2023), https://www.cnbc.com/2023/05/16/googles-palm-2-uses-nearly-five-times-more-text-data-than-predecessor.html.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

but it is also built to generate a wide range of expression from shortform articles to books, chapters, mimicking expressive style, themes of the copyrighted works on which it was trained.

25. Research on LaMDA offers an overview of how the model's 3.6 million tokens were sources to "achieve a more robust performance on dialog tasks:"

a. 50% dialog data from public forums;
b. 12.5% C4 data;
c. 12.5% code documents from sites related to programming…;
d. 12.5% Wikipedia (English);
e. 6.25% English web documents;
f. 6.25% Non-English web documents.[13]

26. Only items (b) "C4 data" and (e) "Wikipedia" are comprised of known data. The remaining 75% of the LaMDA dataset are ambiguous, generalized descriptors for websites and documents found across the internet. As one publication put it, "murky is the best word for describing the 75% of data that Google used for training LaMDA."[14]

27. Of Google's named sources, the C4 dataset contains copyrighted materials, including works found on pirating sites or nonconsenting digital libraries and subscription services. *See, infra* ¶¶ 39-48. While much of Bard/Gemini's training material remains within a proverbial black box, the datasets that developers have made public reveal misappropriation of copyrighted works. *Id.*

28. Google released Gemini publicly on May 10, 2023, in over 180 countries and territories. Gemini quickly reached 142.6 million users the same month.[15] Google plans to expand to more countries, with an anticipated global reach of 1 billion users, or an eighth of all people worldwide.[16] Importantly, Google was determined to expedite the launch of its AI Products at the expense of creators' exclusive rights—secretly harvesting millions of copyrighted materials from the internet without creators' knowledge, consent, and consideration.

29. Google's severe violation of artistic expression harms creators in a number of ways. Most clearly, they lose out on the monetary incentive that comes with being the holder of a

[13] Romal Thopplin, *et al., LaMDA: Language Models for Dialog Applications*, GOOGLE (Feb. 10, 2022), available at: https://arxiv.org/pdf/2201.08239
[14] Roger Montti, *Google Bard AI—What Sites Were Used To Train It*, SEARCH ENGINE JOURNAL (Feb. 10, 2023), https://www.searchenginejournal.com/google-bard-training-data/478941/
[15] *Id.*; David F. Carr, *As ChatGPT Growth Flattened in May, Google Bard Rose 187%*, SIMILARWEB: BLOG (June 5, 2023), https://www.similarweb.com/blog/insights/ai-news/chatgpt-bard/.
[16] Ritik Sharma, *23 Amazing Google Bard Statistics (Users, Facts)*, CONTENTDETECTOR.AI (June 28, 2023), https://contentdetector.ai/articles/google-bard-statistics.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

copyright. Google is not licensing the works as it is required, but instead is stealing them, by reproducing them and integrating them into their AI Products, without the consent of or compensation to the affected creators.

30. But ultimately, this disrespects the very notion of being an author, artist, or creator. These individuals devote a significant amount of time, energy, effort, creativity, and heart to develop their work. To illustrate, writing a book takes years and requires an author to jump through a variety of hurdles. They must complete a manuscript, find an agent, send their manuscript to an editor, and obtain a book deal with a publisher.[17] An author spends months to years to write a book. If an author were zealously writing 500 words a day for seven days a week, it would still take around five and a half months to complete a 300-page book.[18] It can then take anywhere from weeks to several years for an editor to complete their revisions of the book.[19] An author certainly does not undertake this extremely arduous process just for a massive corporation to take it for free.

31. When Google steals and reproduces the works without payment to the authors to make a multi-billion-dollar invention, it deprives authors of deserved royalties, undermining their financial stability and incentives to create new works.

32. Rather than licensing this valuable creative data from the owners, Defendant elected to systematically steal intellectual property of others, including personal and copyright information obtained without consent, and utilize datasets (such as the C-4 dataset) that are riddled with copyrighted and protected works, with the copyright symbol appearing more than 200 million times within the dataset.[20]

33. The law does not allow this kind of systematic infringement that Defendant has been and continues to commit.

---

[17] *How Can I Get Published?* PENGUIN RANDOM HOUSE, https://www.penguinrandomhouse.com/articles/how-can-i-get-published/ (last visited June 26, 2024).

[18] *How Long Does It Take to Write a Book?* MASTERCLASS (Mar. 28, 2022), https://www.masterclass.com/articles/how-long-does-it-take-to-write-a-book.

[19] *How Can I Get Published? Supra* note 17.

[20] Kevin Schual, et. al., *Inside the secret list of websites that make AI like ChatGPT sound smart,* THE WASHINGTON POST (April 19, 2023), https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

34. "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."[21] Scraping is stealing and infringing on the protected intellectual property of others. To do this, Google has to reproduce the works to then use them to build their AI Products. According to a computer science professor at the University of Oxford, the full extent of personal and copyrighted data taken by Defendant's scraping is "unimaginable."[22]

35. Authors, including Plaintiff, publish books with certain copyright management information. This information includes the book's title, the ISBN number or copyright number, the author's name, the copyright holder's name, and terms and conditions of use. Most commonly, this information is found on the back of the book's title page and is customarily included in all books, regardless of genre. The copyright symbol appears on the cover and initial pages of the books. Therefore, when Google downloads, copies, and otherwise reproduces creative works like Plaintiff's book, to train its AI Products, it knows that the works are registered with the copyright office.

36. Google's theft on a massive scale of copyrighted creative works without permission raises serious legal, moral, and ethical questions. Regulators and courts worldwide are seeking to crack down on AI companies "hoovering up content without consent or notice,"[23] but the response by Google and others has been to keep its training datasets largely secret. Google has not permitted any regulatory or other audit access.

///

///

///

[21] Brian Stuenkel, *Personal Information and Artificial Intelligence: Website Scraping and the California Consumer Privacy Act*, COLO. TECH. L. J. (Nov. 2, 2021), https://ctlj.colorado.edu/?p=840.

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

[23] *Id.*

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

A. Google's AI Product Development Depends on Stolen Vast Troves of Copyrighted Data, Including Plaintiff's Book.

37. Around 50% of the C-4 dataset, created by Google in 2020, is taken from the Common Crawl dataset.[24] The Common Crawl dataset is a massive collection of web pages and websites consisting of petabytes of data collected over twelve (12) years, including raw web page data, metadata extracts, and text extracts. The Common Crawl dataset is owned by a non-profit of the same name, which has been indexing and storing as much of the internet as it can access, filing away as many as 3 billion webpages every month, for over a decade.[25]

38. The remaining 50 percent of the C-4 dataset is sourced from what Google vaguely terms as "public forums." The company has declined to clarify the specifics of what constitutes these "public forums," leaving users in the dark about the exact origins and nature of the data influencing half of the AI's training.[26]

39. Importantly, the C-4 dataset is rife with copyrighted and protected works, with the copyright symbol appearing more than 200 million times within the dataset.[27]

40. In fact, the third largest site fueling the dataset is scribd.com, a subscription-based digital library with sixty (60) million e-books and audio books—that compensates authors using a revenue sharing model based on the number of reads their work gets.[28] Word for word excerpts from Plaintiff Leovy's copyrighted work appears on scribd.com. There is no indication Scribd consented to Google's mass misappropriation of copyrighted works on this website, and certainly Plaintiff Leovy nor any other author consented to Google's use of this material, nor were they compensated. Rather, Google has engaged in unauthorized accessing and theft of copyrighted and restricted materials.

---

[24] *Id.*; Katyanna Quach, *4chan and Other Web Sewers Scraped Up Into Google's Mega-Library for Training ML*, THE REGISTER (Apr. 20, 2023), https://www.theregister.com/2023/04/20/google_c4_data_nasty_sources/.

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

[26] Roger Montti, *Google Bard AI: What Sites Were Used to Train It*, SEARCH ENGINE J. (Feb. 10, 2023), https://www.searchenginejournal.com/google-bard-training-data/478941/.

[27] Schaul, *supra* note 20.

[28] *Id.*; Omar, *Scribd Review: Scribd Membership Options, Pros, Cons, and Pricing*, OJ DIGIT. SOLUTIONS (Last updated April 29, 2023), https://ojdigitalsolutions.com/scribd-review/.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

41. Worse, the dataset Google used also contains data from "b-ok.org" a "notorious market for pirated e-books," as well as "[a]t least 27 other sites identified by the U.S. government as markets for piracy and counterfeits."[29]

42. "B-ok.org," also known as "Z Library," is "[t]he world's largest [illegal] ebook library and digital library."[30] It is a pirated cite that is being prosecuted for criminal copyright infringement, and the books that illegally appeared on Z Library, like Plaintiff Leovy's book in its entirety, were not legally authorized to be available for distribution or copying.

43. Google nonetheless took Plaintiff Leovy's book in its entirety, unlawfully misappropriated and reproduced it and utilized it to train Google's AI Products, as it did with countless other author's copyrighted works taken from this and myriad other websites.

44. Additionally, there is a trove of personal blogs represented in the misappropriated data—more than half a million, including the tens of thousands of blogs hosted on Medium, a website especially popular with authors and other content creators. Blogs written on WordPress, Tumbler, Blogspot and Live Journal were also among the materials misappropriated by Google.

45. Google also misappropriated copyrighted information from popular crowdfunding and creative websites, Kickstarter and Patreon, giving Gemini access to thousands of artists' and creators' proprietary marketing materials, "raising concerns [Gemini] may copy this work in suggestions to users."[31]

46. Websites and platforms like Patreon are dedicated to getting creators paid for their exclusive content.[32] Google ignored this and stole creators' exclusive works anyway.

B. Google's Revised Privacy Policy Purports to Give it "Permission" to Take Anything Shared Online to Train and Improve Its AI Products, Including Copyrighted Information.

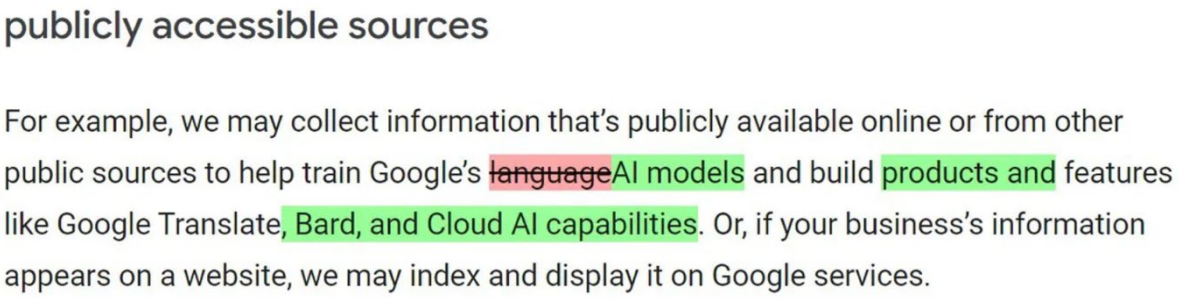
47. On July 1, 2023, Google quietly amended its privacy policy to openly assert that it "collects" publicly available information from the web to train its AI Products, including "Gemini"

[29] Schaul, *supra* note 20.
[30] Beinginstructor, *ZLibrary — The world's largest ebook library*, MEDIUM (Feb. 16, 2023), https://medium.com/@beinginstructor/zlibrary-the-worlds-largest-ebook-library-dfb933762cfc.
[31] Schaul, *supra* note 20.
[32] *The Story of Patreon*, PATREON, https://www.patreon.com/about (last visited June 27, 2024).

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

and "Cloud AI."[33] Given that Google had been doing this theft on a massive scale in secret for years, this disclosure was long overdue. But it was also alarming because it solidified as corporate "policy" Google's disregard for the internet community as a whole, but particularly for authors and creators, and reflected its intent to continue exploiting for commercial gain copyrighted works that are "publicly available online."

Figure 3



48. The idea that Google believes all publicly available creative works on the internet are fair game for it to take, commercially misappropriate, and build AI Products has shocked and angered the creators. As one bestselling author, Alexander Chee, explained, "There's no urgent need to AI to write a novel. The only people who might need that are the people who object to paying writers what they're worth."[34] The CEO of the Author's Guild plainly stated, "It's not fair to use our stuff in your AI without permission or payment. So please start compensating and talking to us."[35]

49. Responding to the backlash, Google announced it will host a public forum to discuss what data collection and protection practices should look like in the new AI era.[36] But as many internet users noted, it is a little too late for that now that Google has already taken and

[33] Jess Weatherbed, *Google Confirms it's Training Bard on Scraped Web Data, Too*, THE VERGE (July 5, 2023), https://www.theverge.com/2023/7/5/23784257/google-ai-bard-privacy-policy-train-web-scraping.
[34] Chloe Veltman, *Thousands of Authors Urge AI Companies to Stop Using Work Without Permission*, NPR (July 17, 2023), https://www.npr.org/2023/07/17/1187523435/thousands-of-authors-urge-ai-companies-to-stop-using-work-without-permission.
[35] *Id.*
[36] Matt G. Southern, *Google Calls for Public Discussion on AI Use of Web Content*, SEARCH ENGINE J. (July 7, 2023), https://www.searchenginejournal.com/google-calls-for-public-discussion-on-ai-use-of-web-content/491053/.

misappropriated copyrighted works. In the words of one, Google is essentially saying to the world: "Now that we've already trained our LLMs on all your proprietary and copyrighted content, we will finally start thinking about giving you a way to opt out of any of your future content being used to make us rich."[37] Google's willingness to *potentially* allow authors to "opt out" of the future use content does not change the fact that Google has illegally copied and misappropriated the copyrighted works, and continues to engage in copyright infringement when it engages in further theft of the same databases and websites that contain pirated works.

50. Defendant's illegal and invasive data misappropriation and infringement practices have also led social platforms that contain copyright protected content, like Twitter and Reddit, to enact more stringent measures in an effort to protect the rights and data of its millions of users.[38] But these anti-scraping modifications stand to negatively impact use of the internet for everyone. For example, now the public cannot view tweets unless they are logged in to Twitter and are limited in how many tweets they can view in one day.

51. These negative impacts to the internet at large underscore the unfortunate ripple effects of Google's misconduct.[39] Unless Google and other AI giants like it are ordered to stop the illegal theft of copyrighted material, other websites might be forced to similarly limit access to the public.

52. As one commentator observed, "should sites really have to wall off their mountains of text so that AI companies can't gobble it up and use it to build AI? That makes no sense."[40] If this were to happen at scale, it would forever change how the internet works, limiting its utility for millions of good faith users who do not want to steal data, but simply engage with it legally in accordance with a site's terms of use and the property interests of the content creators themselves.

53. Moreover, the new policy does not except use of copyrighted (or any other) material

[37] *Id.*
[38] *Musk Says Twitter Will Limit How Many Tweets Users Can Read*, REUTERS (July 1, 2023), https://www.reuters.com/technology/musk-says-twitter-applies-temporary-limit-address-data-scraping-system-2023-07-01/.
[39] Cory Woodroof, *Twitter Users Were Furious After the Website Temporarily Applied a Reading Limit*, USA TODAY (July 1, 2023), https://ftw.usatoday.com/lists/twitter-rate-limit-exceeded-elon-musk-angry-reactions.
[40] Josh Marshall, *Twitter, Musk and the Great AI Land Grab*, TALKING POINTS MEMO (July 6, 2023), https://talkingpointsmemo.com/edblog/twitter-musk-and-the-great-ai-land-grab.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

from being included in its scraped data pool further exposing Google's disregard for intellectual and other property rights while also undermining the policies of various publicly accessible websites, which explicitly prohibit *any* data collection or web scraping for the purpose of training AI models.

54. Now that Google has essentially claimed ownership rights over anything online, there is reason to believe that Google will violate the copyright interests of millions more. Indeed, a massive portion of Defendant's misappropriation to date already includes the unauthorized and widespread theft and copying of copyrighted works extending across a wide spectrum of industries that depend on creative and unique content creation.

55. Instead of competing fairly, Defendant illegally copied the unique works of millions of creators to develop and "train" its AI technology, without consent, credit, or fair compensation. This unauthorized theft and usage of copyrighted content stands in stark violation of creators' exclusive rights under copyright law.

56. Considering the magnitude and scale of the copyright violations to date, along with the likelihood that these violations will continue to increase exponentially, content creators will be dissuaded from investing in the considerable costs of producing unique content in electronic formats. This not only threatens to drastically reshape online accessibility of paid, restricted materials, but also imposes economic harm on a substantial number of content creators.

57. Despite the existence of numerous lawful ways to acquire training data, Defendant purposely elected to bypass the legal routes, opting instead to pillage the internet for copyrighted works. The resulting impact has not only infringed upon the rights of countless creators but has created an environment that ultimately discourages creativity and innovation.

58. The AI Products also dramatically undercuts the commercial market for books and other works already created by radically altering the perceived incentives for anyone to purchase the stolen works going forward. This further harms millions of authors and creators in the form of lost profits and otherwise.

C. Google Uses This Stolen Data to Profit by the Billions.

59. Google's unlawful theft of copyrighted material, at no cost to train its AI technology, has and will continue to unjustly enrich Google. For example, Google announced Gemini, which at

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

the time was called "Bard," on February 6, 2023, and the very next day Alphabet Inc.'s market capitalization increased to 1.37 trillion, reaching 1.62 trillion in June of 2023—its highest market capitalization in the past year.[41]

60. Only a few months after announcing Gemini and in the wake of the AI frenzy, Google co-founders Larry Page and Sergey Brin experienced a combined wealth increase of over $18 billion as the company revealed a revamped AI powered search engine.[42] Page's net worth increased by $9.4 billion to $106.9 billion, while Brin's increased by $8.9 billion to $102.1 billion.[43]

61. This is far from a short-lived AI inspired spike. Google cleverly monetizes its AI Products and fails to meaningfully disclose that Google uses valuable copyrighted material to enhance other Google products and services ***and net billions***.[44] Gemini sweeps in profit for Google on the backs of copyrighted authors and creators without paying any licensing fees for the unauthorized reproduction of their works to train Gemini.

62. Google AI's DeepMind is alone now worth around $55 million,[45] yet the individuals and companies that produced the copyrighted material Google illegally copied and misappropriated 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.

D. Creators are Outraged by Google's Theft-Based Training Model

63. Google has continued to harvest mass amounts of copyrighted material despite an outpour of public outrage. Specifically, creators have recognized and expressed discontent with Google's problematic business model, which allows it to unfairly profit off artists, authors, and content creators, and that forces everyone, whether they want to or not, to contribute to building

[41] *Google Announces Bard, Its Rival to Microsoft-Backed ChatGPT*, FORBES (Feb. 8, 2023), https://www.forbes.com/sites/qai/2023/02/08/google-announces-bard-its-rival-to-microsoft-backed-chatgpt/?sh=29ed0fd93791; *Alphabet Market Cap 2010-2023*, MACROTRENDS, https://www.macrotrends.net/stocks/charts/GOOGL/alphabet/market-cap (last visited June 27, 2024).

[42] Biz Carson, *Google Co-Founders Gain $18 Billion as AI Boost Lifts Stock*, BLOOMBERG (May 12, 2023), https://www.bloomberg.com/news/articles/2023-05-12/google-co-founders-gain-17-billion-as-ai-boost-lifts-stock.

[43] *Id.*

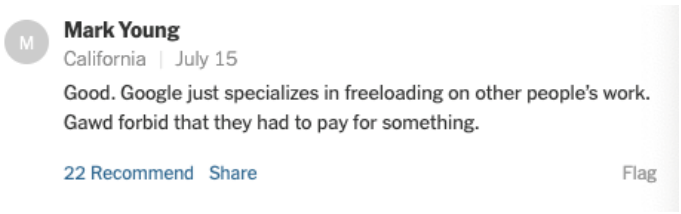
[44] *Bard Privacy Notice*, BARD, https://support.google.com/bard/answer/13594961?hl=en (last updated May 29, 2024).

[45] *DeepMind Net Worth*, PEOPLE AI, https://peopleai.com/fame/identities/deepmind (last visited Jan. 1, 2024).

untested and volatile technology that violates copyright laws.

64. Users are rightfully upset that the content they invest their time and energy into, and, in all cases, which 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, "Authors – your creative work is valuable. It deserves protection. You have the right to control what happens to it. Google is allegedly data scraping all the documents in google docs to train their AI. This includes your work! #writingcommunity."[46]

65. One New York Times reader expressed a similar sentiment: "Google just specializes in freeloading on other people's work. Gawd forbid they had to pay for something."[47]



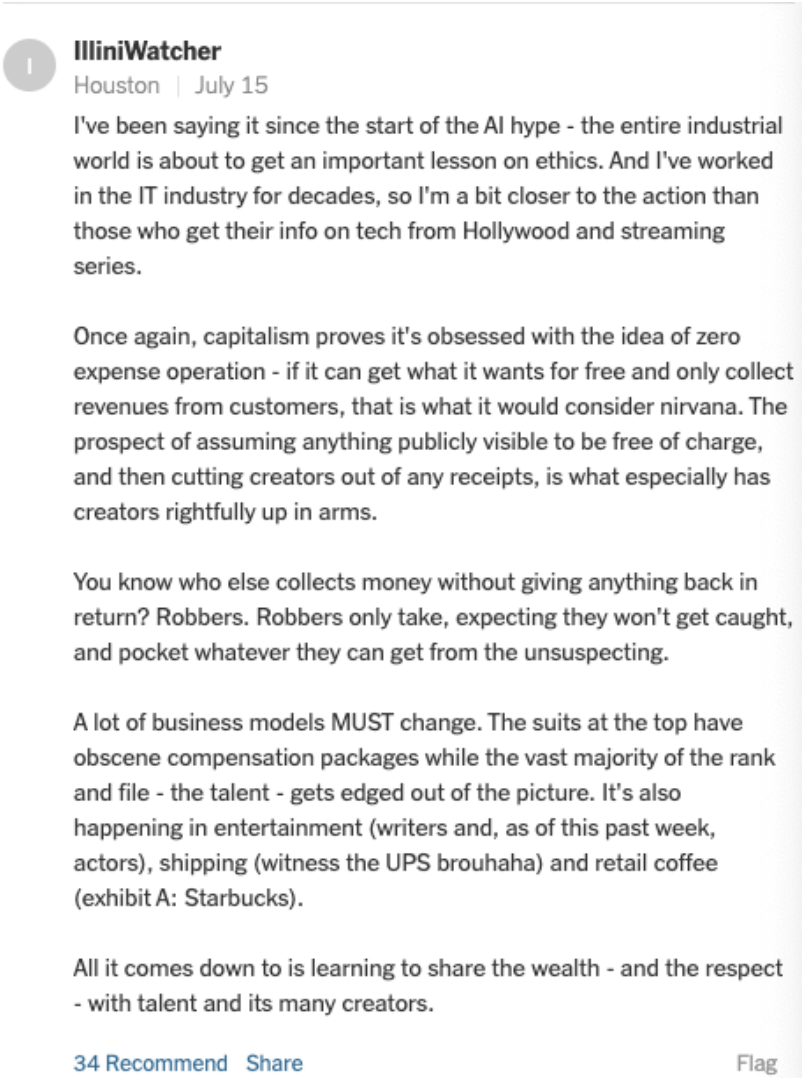
66. Similarly, another New York Times reader added: "Once again, capitalism proves it's obsessed with the idea of a zero-expense operation – if it can get what it wants for free and only

[46] Kelsey Brownlee (@_kelseybrownlee), X (July 14, 2023), https://x.com/_kelseybrownlee/status/1679954300376686594?s=46&t=HHkRbC2AV14Ias3lBERw9g.

[47] 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: Mark Young.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

collect revenues from customers, that is what it could 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."[48] The reader bluntly added, "You know who else collects money without giving anything back in return? Robbers."[49]



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).

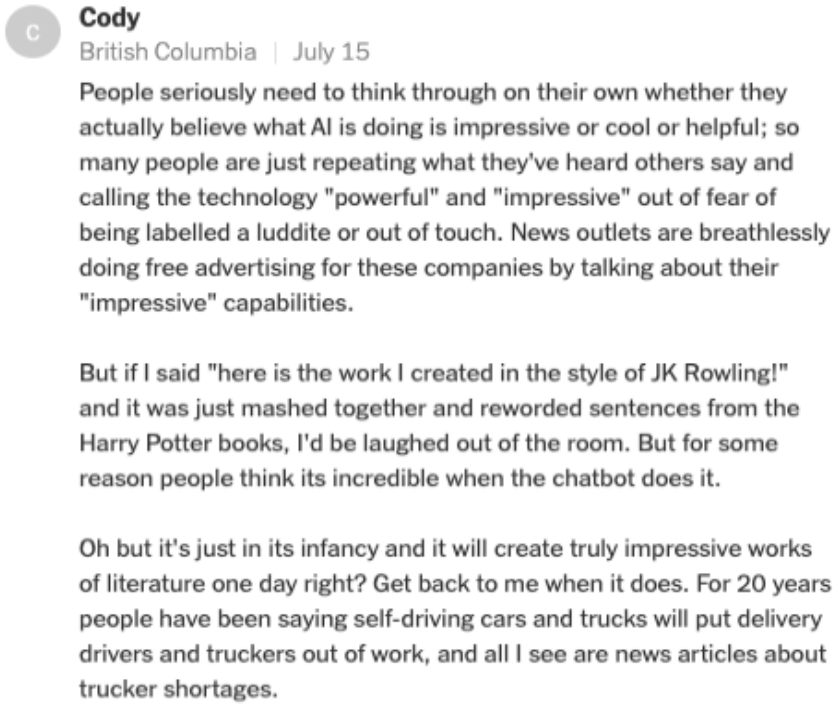
All it comes down to is learning to share the wealth - and the respect - with talent and its many creators.

34 Recommend Share Flag

67. Another reader shared a digestible analogy that proves that users can see through Google'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, I'd be laughed out of the room."[50] Despite AI's smoke-and-mirrors, users can see that big tech's technological advancement is nothing more than wide-scale theft.

[48] *Id.* Commenter: IlliniWatcher.
[49] *Id.*
[50] *Id.* Commenter: Cody.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com



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.

68. 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 stollen, illegally copied, and used to train AI products—without any form of acknowledgement or compensation—they were rightfully upset.

69. In fact, The Author's Guild shared an open letter they wrote to AI companies.[51] 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.[52] Collectively, the authors asked that AI companies, including Google, "Compensate writers fairly for the past and ongoing use of our works in your generative AI programs."[53]

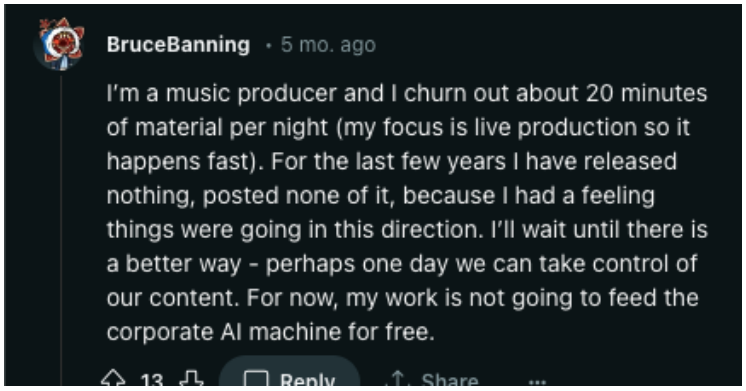
70. Eva Toorenent, an illustrator who serves as the Netherland's advisor for the European Guild for Artificial Intelligence, argued that "[AI models] have sucked the creative juices of millions

[51] The Author's Guild, *Open Letter to Generative AI Leaders*, https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders (last visited June 27, 2024).
[52] *Id.*
[53] *Id.*

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

of artists."[54] 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."[55]

71. The threat of AI companies, like Defendant's, misappropriating and reproducing 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.[56] 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."[57]



72. Absent injunctive relief sought herein, Plaintiff and the Class will continue to not freely contribute online as they might for fear of losing control of their data.

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

73. The US Copyright Office opened a public comment period on August 30, 2023, concerning the use of copyrighted data to train AI models, including the violation of publicity rights.[58]

74. Several individuals noted the glaring invasion of privacy that AI companies are

[54] 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/.
[55] *Id.*
[56] Bruce Banning, *Google's policy update confirms that all your posted content will be utilized for AI training*, REDDIT, (June 2023), https://www.reddit.com/r/technews/comments/14qe9tm/googles_policy_update_confirms_that_all_your/?sort=top.
[57] *Id.*
[58] Emilia David, *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.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

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."[59]

75. 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 scrapping 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 . . . It's violating our trust and privacy, not to mention the amount of copyrighted works it has 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**."[60]

76. The comments exhibited an overwhelming level of infuriation over the sad reality that the creative works 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 everyone is 'having fun' at the expense of my livelihood. Please do not continue letting this companies slide."[61]

F. Online News and Media Businesses are Taking Action Against Google's Unauthorized Infringement

77. Much like the average internet user, many online news and media websites are concerned that Defendant is stealing data to train their AI models.

[59] *Comment from Clorite, Katelyn,* U.S. COPYRIGHT OFFICE (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-1003 (emphasis added).
[60] *Comment from Anonymous*, U.S. COPYRIGHT OFFICE (Oct. 31, 2023), https://www.regulations.gov/comment/COLC-2023-0006-5235.
[61] *Comment from Chan, Maggie*, U.S. COPYRIGHT OFFICE (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-0347.

78. 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 web crawlers, on a go forward basis, are the New York Times, CNN, Reuters, Disney, Bloomberg, The Washington Post, ABC News, ESPN, and Insider.

79. 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.

80. 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. Similar to the reactions of average internet users, these companies' response demonstrates the overarching anger towards Defendant's unfair and anticompetitive practices—spanning across the entire internet food-chain.

II. DEFENDANT'S CONDUCT VIOLATES ESTABLISHED COPYRIGHT LAWS.

81. Defendant's theft and misappropriation violated the copyright rights of all individuals whose creative content was copied and then incorporated into Defendant's AI Products.

82. Defendant's illegal stealing and reproducing of copyrighted works was done largely in secret, without consent from or consideration to any creator whose copyrighted material was reproduced.

83. Creators retain copyright interests over their unique and original content posted (or at times pirated and illegally displayed) online. This content includes text, images, music, video content, and other forms of creative expression, all of which fall under the purview of copyright law.

84. Defendant's unauthorized theft, reproduction, and use of these copyrighted materials constitutes infringement because Defendant copies and downloads the intellectual property to then use it to build and train its AI Products. As an illustrative example, the unauthorized collection and use of copyrighted literary works in training Gemini not only infringes on the rights of the producers

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

but also damages the intrinsic value of the copyrighted works.

85. Copyright protection incentivizes creativity and original content creation. Copyright holders have exclusive rights to reproduce their work in different formats, commercially exploit it, create derivative works, and display or perform the work publicly. Thus, when copyrighted work is co-opted without permission or compensation, as in the case of Defendant's massive theft and infringement, it severely undermines the fundamental principles of copyright law.

86. Further, the practice of illegal theft and infringement of works through web scraping effectively nullifies the concept of "fair use," a critical aspect of copyright law designed to allow limited use of copyrighted material without permission for purposes like commentary, criticism, news reporting, and scholarly reports. *See McGucken v. Pub Ocean Limited*, 42 F.4th 1149 (9th Cir. 2022). Defendant's wholesale collection and use of copyrighted material, with no option for copyright owners to opt out, far exceeds any reasonable interpretation of "fair use." *See VHT v. Zillow Group*, 918 F.3d 723, 743 (9th Cir. 2019); *accord Worldwide Church of God v. Phila. Church of God, Inc.*, 227 F.3d 110, 1118 (9th Cir. 2000) ("[C]opying an entire work militates against a finding of fair use.").

87. The non-consensual aggregation and usage of copyrighted materials disrupts the balance between content creators and consumers that copyright law intends to foster. When original content is unfairly utilized in this manner, it discourages creators from investing time, effort, and resources into creating new content.

88. By using such works as training fodder for its AI, Defendant is not just using these works in an unauthorized manner, but also illegally profiting from infringement. Plaintiff and Class Members have not consented to such exploitation of their copyrighted works. It is only through legal action that the rights of content creators can be protected, and their original works safeguarded against such egregious misuse.

89. While the past, and ongoing, misappropriation of valuable copyrighted material is bad enough, AI Products like Gemini also stand to altogether eliminate future income for millions, due to the widespread unemployment AI and loss of value for intellectual property it expected to cause over time. No one has consented to the use of their copyrighted materials in a manner that not only

violates copyright laws but that also may build this destabilized future of social unrest and worsening poverty for everyday people, while the pockets of Google are lined with profit.

CLASS ALLEGATIONS

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

> **Copyright Class:** All persons in the United States who own a United States copyright in any work that was used as training data for Defendant's Products.

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

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

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

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

95. **Ascertainability:** Membership of the Class is defined based on objective criteria, and individual members will be identifiable from Defendant's records, records of other Google products/services, self-identification methods, or other means. Defendant's records are likely to

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

include massive data storage, user accounts, and data gathered directly from the affected members of Class.

96. **Numerosity:** The precise number of Class Members is not available to Plaintiff, but it is clear that individual joinder is impracticable. Millions of copyright holders have been victims of Defendant's unlawful and unauthorized infringement and theft on massive scale. Class Members can be identified through Defendant's records of copyrighted works, records from the U.S. Copyright Office, or by other means, including but not limited to self-identification.

97. **Commonality:** Commonality requires that the Class Members 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 Class Members are as follows:

a) Whether Leovy owned copyright in her book that was scraped, copied, and used to train Defendant's AI Products.

b) Whether Defendant's conduct constitutes an infringement of the copyrights held by Plaintiff Leovy and the Class in their respective works;

c) Whether Defendant's reproduction of the copyrighted works constitutes a copyright infringement;

d) Whether Defendant's copying and/or reproduction of the copyrighted works constitutes fair use;

e) Whether Defendant's violation of Class' and Plaintiff's exclusive rights under copyright law entitles them to damages, including statutory damages, and the amount of statutory damages;

f) Whether Defendant acted willfully with respect to the copyright infringements;

98. **Typicality:** Plaintiff's claims are typical of the claims of other Class Members in that Plaintiff and the Class Members sustained damages arising out of Defendant's uniform wrongful conduct and data collecting practices, and use of such data in an attempt to train the AI Products, and further develop the Products.

///

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

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

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

a) Implementation of Accountability Protocols that hold Defendant responsible for Products' actions by barring any use of the protected materials until Plaintiff and Class Members are fairly compensated for the stolen and copied protected materials, and are compensated for the ongoing use for their intellectual property;

b) Implementation of Accountability Protocols that hold Defendant responsible for ensuring that during any web scraping it instructs web crawlers to avoid (a) websites/datasets that are known for containing pirated materials (such as Z-library and datasets containing Z-Library); (b) requiring that it reviews copyright notices and terms of the websites/databases which it uses for training to ensure that copyright protected materials are not within the training data sets; (c) limit scraping only to websites/datasets which contain materials within the public domain; (d) limit web scraping to datasets/websites for which Defendant has provided an accepted valuable consideration to authors and creator;

c) Requiring Defendant to allow Product users and everyday internet users to opt out of all collection/copying of their protected works, and stop the illegal taking of protected works, delete any ill-gotten protected materials, or the algorithms which were built on the stolen data;

d) Confirmation that Defendant has deleted, destroyed, and purged the copyrighted materials of all relevant class members unless Defendant can confirm that it has properly licensed such materials; and

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

e) Requiring all further and just corrective action, consistent with permissible law and pursuant to only those causes of action so permitted.

101. **This case also satisfies Fed. R. Civ. P. 23(b)(3) - Predominance:** There are many questions of law and fact common to the claims of Plaintiff and Class Members, 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 Defendant violated Plaintiff's and Class Members' exclusive rights under copyright laws;

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

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

102. **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 Class Members will likely be relatively small, especially given the burden and expense of individual prosecution of the complex litigation necessitated by Defendant's actions. Thus, it would be virtually impossible for the individual Class Members to obtain effective relief from Defendant's 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.

103. 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.

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

COUNT ONE

DIRECT COPYRIGHT INFRINGEMENT

104. Plaintiff Leovy, individually and on behalf of the Class, herein repeats, realleges, and fully incorporates all allegations in all preceding paragraphs.

105. Copyrights are the legal title to intellectual property by which creators of original works (such as books, photographs, videos etc.) protect their moral, economic, and legal rights. The importance of copyrighted works is enshrined in the U.S. Constitution, which expressly gave Congress the power to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." U.S. Const. Art. I, Section 8. "Copyright law encourages people to create original works and thereby 'ultimately serves the purpose of enriching the general public through access to creative works." *Fogerty v. Fantasy, Inc.,* 510 U.S. 517, 526 (1994).

106. The Supreme Court of the United States held that by "establishing a marketable right to the use of one's expression, copyright supplies the economic incentive to create and disseminate ideas." *Harper & Row Publisher, Inc. v. Nation Enters.*, 471 U.S. 539, 558 (1985).

107. The Copyright Act makes it illegal to publicly perform, display, distribute, or reproduce a copyrighted work except in limited instances, and provides for statutory damages, willful statutory damages, and the right to recover attorneys' fees. 17 U.S.C. 501 *et seq.* The Copyright Act grants copyright owners the exclusive public display right, and control of the economic value of their protected works. This is true even where a copyrighted work is displayed somewhere online. Therefore, any person who downloads or even copies the work without consent is infringing on the owner's exclusive rights to reproduction and/or distribution.

108. Defendant copied, downloaded, and otherwise misappropriated a vast trove of protected works available on the internet, including the exact digital version of Plaintiff Leovy's copyrighted book, to develop the Gemini's language model.

109. Defendant's copying, storing processing, and reproducing of the entirety of Plaintiff Leovy's copyrighted book and the copyrighted materials of the Class to train Gemini and other AI Products, infringed on Plaintiff Leovy's and the Class Members' exclusive rights in their

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

copyrighted works. Similarly, Defendant's blatant copying and unlawful appropriation of copyrighted works of others – images, books, song, etc. – infringed on Class Members' exclusive rights.

110. Also, after being trained (by illegally infringing on the copyrighted materials), the AI models are subject to fine turning, wherein Google continues to train/re-train its AI models to better mimic the works, content, expressive style, protected designs, and other important parts of authors' works.

111. Plaintiff Leovy is the exclusive owner of the registered copyright in her work under 17 U.S.C. § 106; in fact, Plaintiff Leovy registered the copyright for her book on February 20, 2015.

112. As exclusive rights holder, only Plaintiff Leovy or those Plaintiff Leovy has authorized may reproduce (i.e. copy, download), or distribute her property. Neither Plaintiff Leovy nor any Class Members authorized Defendant to use their works or make copies of their works.

113. Defendant generates billions of dollars on its AI technology, Gemini, which was trained on copyrighted works and materials without consent or compensation. Without this mass infringement, Gemini would not exist.

114. By training its Products on the protected works of millions of authors, Defendant engaged in unauthorized use, distribution, and reproduction of the copyrighted materials.

115. Defendant's conduct herein is willful because it is aware that stealing works from the entire internet will undoubtedly result in infringement, especially where Defendant is copying the databased and websites that are known to contain pirated books and works. Also, Defendant is aware that the stolen millions of works were registered with the U.S/ Copyright Office, because the copyright symbol appears in at least some datasets used by Defendant (C-4) more than 200 million times.

116. Defendant made copies and engaged in an unauthorized use of Plaintiff Leovy and Class Members' work for training and development of Gemini (as well as other AI Products). Defendant's infringement of copyrighted works was knowing, willful, and intentional, and thus subjects Defendant to liability for statutory damages under Section 504(c)(2) of the Copyright Act of up to $150,000 per infringement. Furthermore, Defendant has sufficient resources to verify

whether or not the works on which Gemini and other AI Products were trained on are protected under copyright law.

117. Alternatively, even if Defendant was unaware and had no reason to believe that its actions constituted copyright infringement, Plaintiff Leovy and Class Members are entitled to $200.00/per infringement.

118. As a direct and proximate cause of Defendant's conduct, Plaintiff Leovy and Class Members have suffered and will continue to suffer monetary damages in an amount to be determined at trial. Plaintiff Leovy and Class Members are entitled to statutory damages, actual damages, disgorgement of profits, injunctive and declaratory relief, and other remedies.

119. Because Plaintiff and the members of the proposed Class have been and continue to be irreparably injured due to Defendant's infringement and conduct described herein, for which no adequate remedy is available at law, Plaintiff and the Class are entitled to injunctive relief. Without permanent injunctive relief, Defendant will continue to infringe on the exclusive rights of Plaintiff and the proposed Class, unless its infringing activity is enjoined by this Court.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff on behalf of herself and the Proposed Class that she seeks 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 Plaintiff to represent the Class;

C. Appoint undersigned counsel to represent the Class;

D. Award compensatory damages to Plaintiff and the Class against Defendant for all damages sustained as a result of Defendant's wrongdoing, in an amount to be proven at trial, including interest;

E. Award statutory (including treble damages, where appropriate) damages to Plaintiff and the Class against Defendant;

F. Award nominal damages to Plaintiff and the Class against Defendant;

G. Non-restitutionary disgorgement of all profits that were derived, in whole or in part,

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com

from Defendant's conduct;

H. Award punitive damages to Plaintiff and the Class against Defendant;

I. Permanently restrain Defendant, 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 Plaintiff and the Class their reasonable costs and expenses incurred in this Action, including attorneys' fees, costs, and expenses; and

K. Grant Plaintiff and the Class such further relief as the Court deems appropriate.

JURY TRIAL DEMANDED

Plaintiff demands a jury trial on all triable issues.

DATED: June 27, 2024

CLARKSON LAW FIRM, P.C.

/s/ Ryan J. Clarkson
Ryan Clarkson, Esq.
Yana Hart, Esq.
Tracey Cowan, Esq.
Tiara Avaness, Esq.

Counsel for Plaintiff and the Proposed Class

Clarkson Law Firm, P.C. | 22525 Pacific Coast Highway, Malibu, CA 90265 | P: (213) 788-4050 F: (213) 788-4070 | clarksonlawfirm.com