Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Kathleen J. McMahon (State Bar No. 340007)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
kmcmahon@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@buttericklaw.com

Counsel for Individual and Representative Plaintiffs and the Proposed Class

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

PAUL TREMBLAY, an individual and MONA AWAD, an individual,

Individual and Representative Plaintiffs,

v.

OPENAI, INC., a Delaware nonprofit corporation; OPENAI, L.P., a Delaware limited partnership; OPENAI OPCO, L.L.C., a Delaware limited liability corporation; OPENAI GP, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND I, L.P., a Delaware limited partnership; and OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company,

Defendants.

Case No.

COMPLAINT

CLASS ACTION

DEMAND FOR JURY TRIAL

Plaintiffs Paul Tremblay and Mona Awad ("Plaintiffs"), on behalf of themselves and all others similarly situated, bring this Class Action Complaint (the "Complaint") against Defendants OpenAI, Inc., OpenAI, L.P., OpenAI OpCo, L.L.C., OpenAI GP, L.L.C., OpenAI Startup Fund I, L.P., OpenAI Startup Fund GP I, L.L.C. and OpenAI Startup Fund Management, LLC for direct copyright infringement, vicarious copyright infringement, violations of section 1202(b) of the Digital Millenium Copyright Act, unjust enrichment, violations of the California and common law unfair competition laws, and negligence. Plaintiffs seek to recover injunctive relief and damages as a result and consequence of Defendants' unlawful conduct.

I. OVERVIEW

1. ChatGPT is a software product created, maintained, and sold by OpenAI.

2. ChatGPT is powered by two AI software programs called GPT-3.5 and GPT-4, also known as *large language models*. Rather than being programmed in the traditional way, a large language model is "trained" by copying massive amounts of text and extracting expressive information from it. This body of text is called the *training dataset*. Once a large language model has copied and ingested the text in its training dataset, it is able to emit convincingly naturalistic text outputs in response to user prompts.

3. A large language model's output is therefore entirely and uniquely reliant on the material in its training dataset. Every time it assembles a text output, the model relies on the information it extracted from its training dataset.

4. Plaintiffs and Class members are authors of books. Plaintiffs and Class members have registered copyrights in the books they published. Plaintiffs and Class members did not consent to the use of their copyrighted books as training material for ChatGPT. Nonetheless, their copyrighted materials were ingested and used to train ChatGPT.

5. Indeed, when ChatGPT is prompted, ChatGPT generates summaries of Plaintiffs' copyrighted works—something only possible if ChatGPT was trained on Plaintiffs' copyrighted works.

6. Defendants, by and through the use of ChatGPT, benefit commercial and profit richly from the use of Plaintiffs' and Class members' copyrighted materials.

II. JURISDICTION AND VENUE

7. 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) and the Digital Millenium Copyright Act (17 U.S.C. § 1202).

8. Jurisdiction and venue is proper in this judicial district under 28 U.S.C. § 1391(c)(2) because defendant OpenAI, Inc. is headquartered in this district, and thus a substantial part of the events giving rise to the claim occurred in this district; and because a substantial part of the events giving rise to Plaintiffs' claims occurred in this District, and a substantial portion of the affected interstate trade and commerce was carried out in this District. 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 has 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.

9. Under Civil Local Rule 3.2(c) and (e), assignment of this case to the San Francisco Division is proper because defendant OpenAI, Inc. is headquartered in San Francisco, a substantial amount part of the events giving rise to Plaintiffs' claims and the interstate trade and commerce involved and affected by Defendants' conduct giving rise to the claims herein occurred in this Division.

III. PARTIES

A. Plaintiffs

10. Plaintiff Paul Tremblay is a writer who lives in Massachusetts. Plaintiff Tremblay owns registered copyrights in several books, including *The Cabin at the End of the World*. This book contains the copyright-management information customarily included in published books, including the name of the author and the year of publication.

11. Plaintiff Mona Awad is a writer who lives in Massachusetts. Plaintiff Awad owns registered copyrights in several books, including *13 Ways of Looking at a Fat Girl* and *Bunny*. These books contain the copyright-management information customarily included in published books, including the name of the author and the year of publication.

12. A nonexhaustive list of registered copyrights owned by Plaintiffs is included as **Exhibit A**.

B. Defendants

13. Defendant OpenAI, Inc. is a Delaware nonprofit corporation with its principal place of business located at 3180 18th St, San Francisco, CA 94110.

14. Defendant OpenAI, L.P. is a Delaware limited partnership with its principal place of business located at 3180 18th St, San Francisco, CA 94110. OpenAI, L.P. is a wholly owned subsidiary of OpenAI Inc. that is operated for profit. OpenAI, Inc. controls OpenAI, L.P. directly and through the other OpenAI entities.

15. Defendant OpenAI OpCo, L.L.C. is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI OpCo, L.L.C. is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI, Inc. controls OpenAI OpCo, L.L.C. directly and through the other OpenAI entities.

16. 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 the general partner of OpenAI, L.P. OpenAI GP manages and operates the day-to-day business and affairs of OpenAI, L.P. OpenAI GP was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI, Inc. directly controls OpenAI GP.

17. 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. OpenAI Startup Fund I was instrumental in the foundation of OpenAI, L.P., including the creation of its business strategy and providing initial funding. OpenAI Startup Fund I was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

18. 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. OpenAI Startup Fund GP I is a party to the unlawful conduct alleged herein. OpenAI Startup Fund GP I manages and operates the day-to-day business and affairs of OpenAI Startup Fund I.

19. 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 is a party to the unlawful conduct alleged herein. OpenAI Startup Fund Management was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

IV. AGENTS AND CO-CONSPIRATORS

20. The unlawful acts alleged against the Defendants in this class action complaint were authorized, ordered, or performed by the Defendants' respective officers, agents, employees, representatives, or shareholders while actively engaged in the management, direction, or control of the Defendants' businesses or affairs. The Defendants' agents operated under the explicit and apparent authority of their principals. Each Defendant, and its subsidiaries, affiliates, and agents operated as a single unified entity.

21. Various persons and/or firms not named as Defendants may have participated as co-conspirators in the violations alleged herein and may have performed acts and made statements in furtherance thereof. Each acted as the principal, agent, or joint venture of, or for other Defendants with respect to the acts, violations, and common course of conduct alleged herein.

V. FACTUAL ALLEGATIONS

22. OpenAI creates and sells artificial-intelligence software products. *Artificial intelligence* is commonly abbreviated "AI." AI software is designed to algorithmically simulate human reasoning or inference, often using statistical methods.

23. Certain AI products created and sold by OpenAI are known as *large language models.* A large language model (or "LLM" for short) is AI software designed to parse and emit natural language. Though a large language model is a software program, it is not created the way most software programs are—that is, by human software engineers writing code. Rather, a large language model is "trained" by copying massive amounts of text from various sources and feeding these copies into the model. This corpus of input material is called the *training dataset*. During training, the large language model copies each piece of text in the training dataset and extracts expressive information from it. The large language model progressively adjusts its output to more closely resemble the sequences of words copied from

the training dataset. Once the large language model has copied and ingested all this text, it is able to emit convincing simulations of natural written language as it appears in the training dataset.

24. Much of the material in OpenAI's training datasets, however, comes from copyrighted works—including books written by Plaintiffs—that were copied by OpenAI without consent, without credit, and without compensation.

25. Authors, including Plaintiffs, 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.

26. OpenAI has released a series of large language models, including GPT-1 (released June 2018), GPT-2 (February 2019), GPT-3 (May 2020), GPT-3.5 (March 2022), and most recently GPT-4 (March 2023). "GPT" is an abbreviation for "generative pre-trained transformer," where *pre-trained* refers to the use of textual material for training, *generative* refers to the model's ability to emit text, and *transformer* refers to the underlying training algorithm. Together, OpenAI's large language models will be referred to as the "OpenAI Language Models."

27. Many kinds of material have been used to train large language models. Books, however, have always been a key ingredient in training datasets for large language models because books offer the best examples of high-quality longform writing.

28. For instance, in its June 2018 paper introducing GPT-1 (called "Improving Language Understanding by Generative Pre-Training"), OpenAI revealed that it trained GPT-1 on BookCorpus, a collection of "over 7,000 unique unpublished books from a variety of genres including Adventure, Fantasy, and Romance." OpenAI confirmed why a dataset of books was so valuable: "Crucially, it contains long stretches of contiguous text, which allows the generative model to learn to condition on long-range information." Hundreds of large language models have been trained on BookCorpus, including those made by OpenAI, Google, Amazon, and others.

29. BookCorpus, however, is a controversial dataset. It was assembled in 2015 by a team of AI researchers for the purpose of training language models. They copied the books from a website

called Smashwords.com that hosts unpublished novels that are available to readers at no cost. Those novels, however, are largely under copyright. They were copied into the BookCorpus dataset without consent, credit, or compensation to the authors.

30. OpenAI also copied many books while training GPT-3. In the July 2020 paper introducing GPT-3 (called "Language Models are Few-Shot Learners"), OpenAI disclosed that 15% of the enormous GPT-3 training dataset came from "two internet-based books corpora" that OpenAI simply called "Books1" and "Books2".

31. Tellingly, OpenAI has never revealed what books are part of the Books1 and Books2 datasets. Though there are some clues. First, OpenAI admitted these are "internet-based books corpora". Second, both Books1 and Books2 are apparently much larger than BookCorpus. Based on numbers given in OpenAI's paper about GPT-3, Books1 is apparently about nine times larger; Books2 is about 42 times larger. Since BookCorpus contained about 7,000 titles, this suggests Books1 would contain about 63,000 titles; Books2 would contain about 294,000 titles.

32. But there are only a handful of "internet-based books corpora" that would be able to deliver this much material.

33. As noted in Paragraph 31, *supra*, the OpenAI Books1 dataset can be estimated to contain about 63,000 titles. Project Gutenberg is an online archive of e-books whose copyright has expired. In September 2020, Project Gutenberg claimed to have "over 60,000" titles. Project Gutenberg has long been popular for training AI systems due to the lack of copyright. In 2018, a team of AI researchers created the "Standardized Project Gutenberg Corpus", which contained "more than 50,000 books". On information and belief, the OpenAI Books1 dataset is based on either the Standardized Project Gutenberg Corpus or Project Gutenberg itself, because of the roughly similar sizes of the two datasets.

34. As noted in Paragraph 31, *supra*, the OpenAI Books2 dataset can be estimated to 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. These flagrantly illegal shadow libraries have long been of interest to the AI-training community: for instance, 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. On information and belief, the OpenAI Books2 dataset includes books copied from these "shadow libraries", because those are the most sources of trainable books most similar in nature and size to OpenAI's description of Books2.

35. In March 2023, OpenAI's paper introducing GPT-4 contained no information about its dataset at all: OpenAI claimed that "[g]iven both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about . . . dataset construction." Later in the paper, OpenAI concedes it did "filter[ ] our dataset . . . to specifically reduce the quantity of inappropriate erotic text content."

A. Interrogating the OpenAI Language Models using ChatGPT

36. ChatGPT is a language model created and sold by OpenAI. As its name suggests, ChatGPT is designed to offer a conversational style of interaction with a user. OpenAI offers ChatGPT through a web interface to individual users for $20 per month. Through the web interface, users can choose to use two versions of ChatGPT: one based on the GPT-3.5 model, and one based on the newer GPT-4 model.

37. OpenAI also offers ChatGPT to software developers through an application-programming interface (or "API"). The API allows developers to write programs that exchange data with ChatGPT. Access to ChatGPT via the API is billed on the basis of usage.

38. Regardless of how accessed—either through the web interface or through the API—ChatGPT allows users to enter text prompts, which ChatGPT then attempts to respond to in a natural way, i.e., ChatGPT can generate answers in a coherent and fluent way that closely mimics human language. If a user prompts ChatGPT with a question, ChatGPT will answer. If a user prompts ChatGPT with a command, ChatGPT will obey. If a user prompts ChatGPT to summarize a copyrighted book, it will do so.

39. ChatGPT's output, like other LLMs, relies on the data upon which it is trained to generate new content. LLMs generate output based on patterns and connections drawn from the training data. For example, if an LLM is prompted to generate a writing in the style of a certain author,

the LLM would generate content based on patterns and connections it learned from analysis of that author's work within its training data.

40. On information and belief, the reason ChatGPT can accurately summarize a certain copyrighted book is because that book was copied by OpenAI and ingested by the underlying OpenAI Language Model (either GPT-3.5 or GPT-4) as part of its training data.

41. When ChatGPT was prompted to summarize books written by each of the Plaintiffs, it generated very accurate summaries. These summaries are attached as **Exhibit B**. The summaries get some details wrong. These details are highlighted in the summaries. This is expected, since a large language model mixes together expressive material derived from many sources. Still, the rest of the summaries are accurate, which means that ChatGPT retains knowledge of particular works in the training dataset and is able to output similar textual content. At no point did ChatGPT reproduce any of the copyright management information Plaintiffs included with their published works.

VI. CLASS ALLEGATIONS

A. Class Definition

42. Plaintiffs bring this action for damages and injunctive relief as a class action under Federal Rules of Civil Procedure 23(a), 23(b)(2), and 23(b)(3), on behalf of the following Class:

> **All persons or entities domiciled in the United States that own a United States copyright in any work that was used as training data for the OpenAI Language Models during the Class Period.**

43. This Class definition excludes:

a. any of the Defendants named herein;

b. any of the Defendants' co-conspirators;

c. any of Defendants' parent companies, subsidiaries, and affiliates;

d. any of Defendants' officers, directors, management, employees, subsidiaries, affiliates, or agents;

e. all governmental entities; and

f. the judges and chambers staff in this case, as well as any members of their immediate families.

B. Numerosity

44. Plaintiffs do not know the exact number of members in the Class. This information is in the exclusive control of Defendants. On information and belief, there are at least thousands of members in the Class geographically dispersed throughout the United States. Therefore, joinder of all members of the Class in the prosecution of this action is impracticable.

C. Typicality

45. Plaintiffs' claims are typical of the claims of other members of the Class because Plaintiffs and all members of the Class were damaged by the same wrongful conduct of Defendants as alleged herein, and the relief sought herein is common to all members of the Class.

D. Adequacy

46. Plaintiffs will fairly and adequately represent the interests of the members of the Class because the Plaintiffs have experienced the same harms as the members of the Class and have no conflicts with any other members of the Class. Furthermore, Plaintiffs have retained sophisticated and competent counsel who are experienced in prosecuting federal and state class actions, as well as other complex litigation.

E. Commonality and Predominance

47. Numerous questions of law or fact common to each Class arise from Defendants' conduct:

a. whether Defendants violated the copyrights of Plaintiffs and the Class when they downloaded copies of Plaintiff's copyrighted books and used them to train ChatGPT;

b. whether ChatGPT itself is an infringing derivative work based on Plaintiffs' copyrighted books;

c. whether the text outputs of ChatGPT are infringing derivative works based on Plaintiffs' copyrighted books;

d. whether Defendants violated the DMCA by removing copyright-management information (CMI) from Plaintiffs' copyrighted books.

e. Whether Defendants were unjustly enriched by the unlawful conduct alleged herein.

f. Whether Defendants' conduct alleged herein constitutes Unfair Competition under

California Business and Professions Code section 17200 *et seq*.

g. Whether Defendants' conduct alleged herein constitutes unfair competition under the common law.

h. Whether this Court should enjoin Defendants from engaging in the unlawful conduct alleged herein. And what the scope of that injunction would be.

i. Whether any affirmative defense excuses Defendants' conduct.

j. Whether any statutes of limitation limits Plaintiffs' and the Class's potential for recovery.

48. These and other questions of law and fact are common to the Class predominate over any questions affecting the members of the Class individually.

F. Other Class Considerations

49. Defendants have acted on grounds generally applicable to the Class. This class action is superior to alternatives, if any, for the fair and efficient adjudication of this controversy. Prosecuting the claims pleaded herein as a class action will eliminate the possibility of repetitive litigation. There will be no material difficulty in the management of this action as a class action. Further, final injunctive relief is appropriate with respect to the Class as a whole.

50. The prosecution of separate actions by individual Class members would create the risk of inconsistent or varying adjudications, establishing incompatible standards of conduct for Defendants.

VII. CLAIMS FOR RELIEF

COUNT I
Direct Copyright Infringement
17 U.S.C. § 106
On Behalf of Plaintiffs and the Class

51. Plaintiffs incorporate by reference the preceding factual allegations.

52. As the owners of the registered copyrights in books used to train the OpenAI Language Models, Plaintiffs hold the exclusive rights to those texts under 17 U.S.C. § 106.

53. Plaintiffs never authorized OpenAI to make copies of their books, make derivative works, publicly display copies (or derivative works), or distribute copies (or derivative works). All those rights belong exclusively to Plaintiffs under copyright law.

54. On information and belief, to train the OpenAI Language Models, OpenAI relied on harvesting mass quantities of textual material from the public internet, including Plaintiffs' books, which are available in digital formats.

55. OpenAI made copies of Plaintiffs' books during the training process of the OpenAI Language Models without Plaintiffs' permission. Specifically, OpenAI copied at least Plaintiff Tremblay's book *The Cabin at the End of the World*; and Plaintiff Awad's books *13 Ways of Looking at a Fat Girl* and *Bunny*. Together, these books are referred to as the **Infringed Works**.

56. Because the OpenAI Language Models cannot function without the expressive information extracted from Plaintiffs' works (and others) and retained inside them, the OpenAI Language Models are themselves infringing derivative works, made without Plaintiffs' permission and in violation of their exclusive rights under the Copyright Act.

57. Plaintiffs have been injured by OpenAI's acts of direct copyright infringement. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits, and other remedies provided by law.

COUNT 2
Vicarious Copyright Infringement
17 U.S.C. § 106
On Behalf of Plaintiffs and the Class

58. Plaintiffs incorporate by reference the preceding factual allegations.

59. Because the output of the OpenAI Language Models is based on expressive information extracted from Plaintiffs' works (and others), every output of the OpenAI Language Models is an infringing derivative work, made without Plaintiffs' permission and in violation of their exclusive rights under the Copyright Act.

60. OpenAI has the right and ability to control the output of the OpenAI Language Models. OpenAI has benefited financially from the infringing output of the OpenAI Language Models. Therefore, every output from the OpenAI Language Models constitutes an act of vicarious copyright infringement.

61. Plaintiffs have been injured by OpenAI's acts of vicarious copyright infringement. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits, and other remedies provided by law.

COUNT 3
Digital Millenium Copyright Act—Removal of Copyright Management Information
17 U.S.C. § 1202(b)
On Behalf of Plaintiffs and the Class

62. Plaintiffs incorporate by reference the preceding factual allegations.

63. Plaintiffs included one or more forms of copyright-management information ("CMI") in each of the Plaintiffs' Infringed Works, including: copyright notice, title and other identifying information, the name or other identifying information about the owners of each book, terms and conditions of use, and identifying numbers or symbols referring to CMI.

64. Without the authority of Plaintiffs and the Class, OpenAI copied the Plaintiffs' Infringed Works and used them as training data for the OpenAI Language Models. By design, the training process does not preserve any CMI. Therefore, OpenAI intentionally removed CMI from the Plaintiffs' Infringed Works in violation of 17 U.S.C. § 1202(b)(1).

65. Without the authority of Plaintiffs and the Class, Defendants created derivative works based on Plaintiffs' Infringed Works. By distributing these works without their CMI, OpenAI violated 17 U.S.C. § 1202(b)(3).

66. OpenAI knew or had reasonable grounds to know that this removal of CMI would facilitate copyright infringement by concealing the fact that every output from the OpenAI Language Models is an infringing derivative work, synthesized entirely from expressive information found in the training data.

67. Plaintiffs have been injured by OpenAI's removal of CMI. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits, and other remedies provided by law.

COUNT 4
Unfair Competition
Cal. Bus. & Prof. Code §§ 17200, *et seq.*
On Behalf of Plaintiffs and the Class

68. Plaintiffs incorporate by reference the preceding factual allegations.

69. Defendants have engaged in unlawful business practices, including violating Plaintiffs' rights under the DMCA, and using Plaintiffs' Infringed Works to train ChatGPT without Plaintiffs' or the Class's authorization.

70. The unlawful business practices described herein violate California Business and Professions Code section 17200 *et seq*. (the "UCL") because that conduct is otherwise unlawful by violating the DMCA.

71. The unlawful business practices described herein violate the UCL because they are unfair, immoral, unethical, oppressive, unscrupulous or injurious to consumers, because, among other reasons, Defendants used Plaintiffs' protected works to train ChatGPT for Defendants' own commercial profit without Plaintiffs' and the Class's authorization. Defendants further knowingly designed ChatGPT to output portions or summaries of Plaintiffs' copyrighted works without attribution, and they unfairly profit from and take credit for developing a commercial product based on unattributed reproductions of those stolen writing and ideas.

72. The unlawful business practices described herein violate the UCL because consumers are likely to be deceived. Defendants knowingly and secretively trained ChatGPT on unauthorized copies of Plaintiffs' copyright-protected work. Further Defendants deceptively designed ChatGPT to output without any CMI or other credit to Plaintiffs and Class members whose Infringed Works comprise ChatGPT's training dataset. Defendants deceptively marketed their product in a manner that fails to attribute the success of their product to the copyright-protected work on which it is based.

COUNT 5
Negligence
On Behalf of Plaintiffs and the Class

73. Plaintiffs incorporate by reference the preceding factual allegations.

74. Defendants owed a duty of care toward Plaintiffs and the Class based upon Defendants' relationship to them. This duty is based upon Defendants' obligations, custom and practice, right to control information in its possession, exercise of control over the information in its possession, authority to control the information in its possession, and the commission of affirmative acts that result in said harms and losses. Additionally, this duty is based on the requirements of California Civil Code

section 1714, requiring all "persons," including Defendants, to act in a reasonable manner toward others.

75. Defendants breached their duties by negligently, carelessly, and recklessly collecting, maintaining and controlling Plaintiffs' and Class members' Infringed Works and engineering, designing, maintaining and controlling systems—including ChatGPT—which are trained on Plaintiffs' and Class members' Infringed Works without their authorization.

76. Defendants owed Plaintiffs and Class members a duty of care to maintain Plaintiffs' Infringed Works once collected and ingested for training ChatGPT.

77. Defendants also owed Plaintiffs and Class members a duty of care to not use the Infringed Works in a way that would foreseeably cause Plaintiffs and Class members injury, for instance, by using the Infringed Works to train ChatGPT.

78. Defendants breached their duties by, *inter alia*, using Plaintiffs' Infringed Works to train ChatGPT.

COUNT 6
Unjust Enrichment
On Behalf of Plaintiffs and the Class

79. Plaintiffs incorporate by reference the preceding factual allegations.

80. Plaintiffs and the Class have invested substantial time and energy in creating the Infringed Works.

81. Defendants have unjustly utilized access to the Infringed Materials to train ChatGPT.

82. Plaintiffs did not consent to the unauthorized use of the Infringed Materials to train ChatGPT.

83. By using Plaintiffs' Infringed Works to train ChatGPT, Plaintiffs and the Class were deprived of the benefits of their work, including monetary damages.

84. Defendants derived profit and other benefits from the use of the Infringed Materials to train ChatGPT.

85. It would be unjust for Defendants to retain those benefits.

86. The conduct of Defendants is causing and, unless enjoined and restrained by this Court,

will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

VIII. DEMAND FOR JUDGMENT

WHEREFORE, Plaintiffs request that the Court enter judgment on their behalf and on behalf of the Class defined herein, by ordering:

a) This action may proceed as a class action, with Plaintiffs serving as Class Representatives, and with Plaintiffs' counsel as Class Counsel.

b) Judgment in favor of Plaintiffs and the Class and against Defendants.

c) An award of statutory and other damages under 17 U.S.C. § 504 for violations of the copyrights of Plaintiffs and the Class by Defendants.

d) Permanent injunctive relief, including but not limited to changes to ChatGPT to ensure that all applicable information set forth in 17 U.S.C. § 1203(b)(1) is included when appropriate.

e) An order of costs and allowable attorney's fees under 17 U.S.C. § 1203(b)(4)–(5).

f) An award of statutory damages under 17 U.S.C. § 1203(b)(3) and 17 U.S.C. § 1203(c)(3), or in the alternative, an award of actual damages and any additional profits under 17 U.S.C. § 1203(c)(2) (including tripling damages under 17 U.S.C. § 1203(c)(4) if applicable).

g) Pre- and post-judgment interest on the damages awarded to Plaintiffs and the Class, and that such interest be awarded at the highest legal rate from and after the date this class action complaint is first served on Defendants.

h) Defendants are to be jointly and severally responsible financially for the costs and expenses of a Court approved notice program through post and media designed to give immediate notification to the Class.

i) Further relief for Plaintiffs and the Class as may be just and proper.

IX. JURY TRIAL DEMANDED

Under Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury of all the claims asserted in this Complaint so triable.

Dated: June 28, 2023

By: /s/ *Joseph R. Saveri*
Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Kathleen J. McMahon (State Bar No. 340007)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, California 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
kmcmahon@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@buttericklaw.com

Counsel for Individual and Representative Plaintiffs and the Proposed Class