

1 Joseph R. Saveri (State Bar No. 130064)
2 Cadio Zirpoli (State Bar No. 179108)
3 Christopher K.L. Young (State Bar No. 318371)
4 Kathleen J. McMahon (State Bar No. 340007)

JOSEPH SAVERI LAW FIRM, LLP

5 601 California Street, Suite 1000
6 San Francisco, California 94108
7 Telephone: (415) 500-6800
8 Facsimile: (415) 395-9940
9 Email: jsaveri@saverilawfirm.com
10 czirpoli@saverilawfirm.com
11 cyoung@saverilawfirm.com
12 kmcmahon@saverilawfirm.com

13 Matthew Butterick (State Bar No. 250953)
14 1920 Hillhurst Avenue, #406
15 Los Angeles, CA 90027
16 Telephone: (323) 968-2632
17 Facsimile: (415) 395-9940
18 Email: mb@buttericklaw.com

19 *Counsel for Individual and Representative Plaintiffs*
20 *and the Proposed Class*

21 **UNITED STATES DISTRICT COURT**
22 **NORTHERN DISTRICT OF CALIFORNIA**
23 **SAN FRANCISCO DIVISION**

24 PAUL TREMBLAY, an individual and
25 MONA AWAD, an individual,
26
27 Individual and Representative Plaintiffs,

28 v.

OPENAI, INC., a Delaware nonprofit corporation; OPENAI,
L.P., a Delaware limited partnership; OPENAI OP CO, 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

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

9 I. OVERVIEW

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

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

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

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

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

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

28

1 **II. JURISDICTION AND VENUE**

2 7. This Court has subject matter jurisdiction under 28 U.S.C. § 1331 because this case
3 arises under the Copyright Act (17 U.S.C. § 501) and the Digital Millenium Copyright Act (17 U.S.C. §
4 1202).

5 8. Jurisdiction and venue is proper in this judicial district under 28 U.S.C. § 1391(c)(2)
6 because defendant OpenAI, Inc. is headquartered in this district, and thus a substantial part of the
7 events giving rise to the claim occurred in this district; and because a substantial part of the events
8 giving rise to Plaintiffs’ claims occurred in this District, and a substantial portion of the affected
9 interstate trade and commerce was carried out in this District. Each Defendant has transacted business,
10 maintained substantial contacts, and/or committed overt acts in furtherance of the illegal scheme and
11 conspiracy throughout the United States, including in this District. Defendants’ conduct has had the
12 intended and foreseeable effect of causing injury to persons residing in, located in, or doing business
13 throughout the United States, including in this District.

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

18 **III. PARTIES**

19 **A. Plaintiffs**

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

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

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

2 **Exhibit A.**

3 **B. Defendants**

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

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

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

14 16. Defendant OpenAI GP, L.L.C. (“OpenAI GP”) is a Delaware limited liability company
15 with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI GP is
16 the general partner of OpenAI, L.P. OpenAI GP manages and operates the day-to-day business and
17 affairs of OpenAI, L.P. OpenAI GP was aware of the unlawful conduct alleged herein and exercised
18 control over OpenAI, L.P. throughout the Class Period. OpenAI, Inc. directly controls OpenAI GP.

19 17. Defendant OpenAI Startup Fund I, L.P. (“OpenAI Startup Fund I”) is a Delaware
20 limited partnership with its principal place of business located at 3180 18th Street, San Francisco, CA
21 94110. OpenAI Startup Fund I was instrumental in the foundation of OpenAI, L.P., including the
22 creation of its business strategy and providing initial funding. OpenAI Startup Fund I was aware of the
23 unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

24 18. Defendant OpenAI Startup Fund GP I, L.L.C. (“OpenAI Startup Fund GP I”) is a
25 Delaware limited liability company with its principal place of business located at 3180 18th Street, San
26 Francisco, CA 94110. OpenAI Startup Fund GP I is the general partner of OpenAI Startup Fund I.
27 OpenAI Startup Fund GP I is a party to the unlawful conduct alleged herein. OpenAI Startup Fund GP
28 I manages and operates the day-to-day business and affairs of OpenAI Startup Fund I.

1 19. Defendant OpenAI Startup Fund Management, LLC (“OpenAI Startup Fund
2 Management”) is a Delaware limited liability company with its principal place of business located at
3 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund Management is a party to the
4 unlawful conduct alleged herein. OpenAI Startup Fund Management was aware of the unlawful
5 conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

6 IV. AGENTS AND CO-CONSPIRATORS

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

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

17 V. FACTUAL ALLEGATIONS

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

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

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

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

6 25. Authors, including Plaintiffs, publish books with certain copyright management
7 information. This information includes the book’s title, the ISBN number or copyright number, the
8 author’s name, the copyright holder’s name, and terms and conditions of use. Most commonly, this
9 information is found on the back of the book’s title page and is customarily included in all books,
10 regardless of genre.

11 26. OpenAI has released a series of large language models, including GPT-1 (released June
12 2018), GPT-2 (February 2019), GPT-3 (May 2020), GPT-3.5 (March 2022), and most recently GPT-4
13 (March 2023). “GPT” is an abbreviation for “generative pre-trained transformer,” where *pre-trained*
14 refers to the use of textual material for training, *generative* refers to the model’s ability to emit text, and
15 *transformer* refers to the underlying training algorithm. Together, OpenAI’s large language models will
16 be referred to as the “OpenAI Language Models.”

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

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

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

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

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

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

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

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

23 34. As noted in Paragraph 31, *supra*, the OpenAI Books2 dataset can be estimated to contain
24 about 294,000 titles. The only “internet-based books corpora” that have ever offered that much
25 material are notorious “shadow library” websites like Library Genesis (aka LibGen), Z-Library (aka B-
26 ok), Sci-Hub, and Bibliotik. The books aggregated by these websites have also been available in bulk via
27 torrent systems. These flagrantly illegal shadow libraries have long been of interest to the AI-training
28 community: for instance, an AI training dataset published in December 2020 by EleutherAI called

1 “Books3” includes a recreation of the Bibliotik collection and contains nearly 200,000 books. On
2 information and belief, the OpenAI Books2 dataset includes books copied from these “shadow
3 libraries”, because those are the most sources of trainable books most similar in nature and size to
4 OpenAI’s description of Books2.

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

10 **A. Interrogating the OpenAI Language Models using ChatGPT**

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

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

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

25 39. ChatGPT’s output, like other LLMs, relies on the data upon which it is trained to
26 generate new content. LLMs generate output based on patterns and connections drawn from the
27 training data. For example, if an LLM is prompted to generate a writing in the style of a certain author,
28

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

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

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

13 VI. CLASS ALLEGATIONS

14 A. Class Definition

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

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

20 43. This Class definition excludes:

- 21 a. any of the Defendants named herein;
- 22 b. any of the Defendants' co-conspirators;
- 23 c. any of Defendants' parent companies, subsidiaries, and affiliates;
- 24 d. any of Defendants' officers, directors, management, employees, subsidiaries,
25 affiliates, or agents;
- 26 e. all governmental entities; and
- 27 f. the judges and chambers staff in this case, as well as any members of their
28 immediate families.

1 **B. Numerosity**

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

6 **C. Typicality**

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

10 **D. Adequacy**

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

16 **E. Commonality and Predominance**

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

- 19 a. whether Defendants violated the copyrights of Plaintiffs and the Class when they
20 downloaded copies of Plaintiff's copyrighted books and used them to train ChatGPT;
21 b. whether ChatGPT itself is an infringing derivative work based on Plaintiffs' copyrighted
22 books;
23 c. whether the text outputs of ChatGPT are infringing derivative works based on Plaintiffs'
24 copyrighted books;
25 d. whether Defendants violated the DMCA by removing copyright-management
26 information (CMI) from Plaintiffs' copyrighted books.
27 e. Whether Defendants were unjustly enriched by the unlawful conduct alleged herein.
28 f. Whether Defendants' conduct alleged herein constitutes Unfair Competition under

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

- 2 g. Whether Defendants' conduct alleged herein constitutes unfair competition under the
3 common law.
- 4 h. Whether this Court should enjoin Defendants from engaging in the unlawful conduct
5 alleged herein. And what the scope of that injunction would be.
- 6 i. Whether any affirmative defense excuses Defendants' conduct.
- 7 j. Whether any statutes of limitation limits Plaintiffs' and the Class's potential for recovery.

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

10 **F. Other Class Considerations**

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

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

19 **VII. CLAIMS FOR RELIEF**

20 **COUNT I**
21 **Direct Copyright Infringement**
22 **17 U.S.C. § 106**
On Behalf of Plaintiffs and the Class

23 51. Plaintiffs incorporate by reference the preceding factual allegations.

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

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

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

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

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

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

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

18 58. Plaintiffs incorporate by reference the preceding factual allegations.

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

23 60. OpenAI has the right and ability to control the output of the OpenAI Language Models.
 24 OpenAI has benefited financially from the infringing output of the OpenAI Language Models.
 25 Therefore, every output from the OpenAI Language Models constitutes an act of vicarious copyright
 26 infringement.
 27
 28

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

4 **COUNT 3**
5 **Digital Millenium Copyright Act—Removal of Copyright Management Information**
6 **17 U.S.C. § 1202(b)**
7 **On Behalf of Plaintiffs and the Class**

8 62. Plaintiffs incorporate by reference the preceding factual allegations.

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

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

17 65. Without the authority of Plaintiffs and the Class, Defendants created derivative works
18 based on Plaintiffs’ Infringed Works. By distributing these works without their CMI, OpenAI violated
19 17 U.S.C. § 1202(b)(3).

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

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

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

68. Plaintiffs incorporate by reference the preceding factual allegations.

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

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

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

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

20 **COUNT 5**
21 **Negligence**
22 **On Behalf of Plaintiffs and the Class**

23 73. Plaintiffs incorporate by reference the preceding factual allegations.

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

1 section 1714, requiring all “persons,” including Defendants, to act in a reasonable manner toward
2 others.

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

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

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

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

14 **COUNT 6**
15 **Unjust Enrichment**
16 **On Behalf of Plaintiffs and the Class**

17 79. Plaintiffs incorporate by reference the preceding factual allegations.

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

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

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

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

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

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

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

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

3 **VIII. DEMAND FOR JUDGMENT**

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

- 6 a) This action may proceed as a class action, with Plaintiffs serving as Class
7 Representatives, and with Plaintiffs' counsel as Class Counsel.
- 8 b) Judgment in favor of Plaintiffs and the Class and against Defendants.
- 9 c) An award of statutory and other damages under 17 U.S.C. § 504 for violations of the
10 copyrights of Plaintiffs and the Class by Defendants.
- 11 d) Permanent injunctive relief, including but not limited to changes to ChatGPT to ensure
12 that all applicable information set forth in 17 U.S.C. § 1203(b)(1) is included when
13 appropriate.
- 14 e) An order of costs and allowable attorney's fees under 17 U.S.C. § 1203(b)(4)–(5).
- 15 f) An award of statutory damages under 17 U.S.C. § 1203(b)(3) and 17 U.S.C. § 1203(c)(3),
16 or in the alternative, an award of actual damages and any additional profits under 17
17 U.S.C. § 1203(c)(2) (including tripling damages under 17 U.S.C. § 1203(c)(4) if
18 applicable).
- 19 g) Pre- and post-judgment interest on the damages awarded to Plaintiffs and the Class, and
20 that such interest be awarded at the highest legal rate from and after the date this class
21 action complaint is first served on Defendants.
- 22 h) Defendants are to be jointly and severally responsible financially for the costs and
23 expenses of a Court approved notice program through post and media designed to give
24 immediate notification to the Class.
- 25 i) Further relief for Plaintiffs and the Class as may be just and proper.

26 **IX. JURY TRIAL DEMANDED**

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

1 Dated: June 28, 2023

2 By: /s/ Joseph R. Saveri
3 Joseph R. Saveri

4 Joseph R. Saveri (State Bar No. 130064)
5 Cadio Zirpoli (State Bar No. 179108)
6 Christopher K.L. Young (State Bar No. 318371)
7 Kathleen J. McMahon (State Bar No. 340007)
8 **JOSEPH SAVERI LAW FIRM, LLP**
9 601 California Street, Suite 1000
10 San Francisco, California 94108
11 Telephone: (415) 500-6800
12 Facsimile: (415) 395-9940
13 Email: jsaveri@saverilawfirm.com
14 czirpoli@saverilawfirm.com
15 cyoung@saverilawfirm.com
16 kmcmahon@saverilawfirm.com

17 Matthew Butterick (State Bar No. 250953)
18 1920 Hillhurst Avenue, #406
19 Los Angeles, CA 90027
20 Telephone: (323) 968-2632
21 Facsimile: (415) 395-9940
22 Email: mb@buttericklaw.com

23 *Counsel for Individual and Representative*
24 *Plaintiffs and the Proposed Class*