

**JOSEPH SAVERI LAW FIRM, LLP**  
Joseph R. Saveri (SBN 130064)  
601 California Street, Suite 1505  
San Francisco, CA 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940  
jsaveri@saverilawfirm.com

**BOIES SCHILLER FLEXNER LLP**  
David Boies (*pro hac vice*)  
333 Main Street  
Armonk, NY 10504  
(914) 749-8200  
dboies@bsflp.com

**CAFFERTY CLOBES MERIWETHER  
& SPRENGEL LLP**  
Bryan L. Clobes (*pro hac vice*)  
135 South LaSalle Street, Suite 3210  
Chicago, IL 60603  
Tel: (312) 782-4880  
bclobes@caffertyclobes.com

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

[Additional counsel on signature page]

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

CATHERINE DENIAL an individual, IAN  
MCDOWELL, an individual, AND STEVEN  
SCHWARTZ, an individual

v.

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., OPENAI STARTUP FUND  
MANAGEMENT, LLC., and MICROSOFT  
CORPORATION

**COMPLAINT**

**Class Action**

**Demand For Jury Trial**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

**TABLE OF CONTENTS**

I. INTRODUCTION .....1  
II. OVERVIEW.....2  
III. JURISDICTION AND VENUE.....4  
IV. PARTIES .....4  
V. FACTUAL ALLEGATIONS.....6  
VI. MICROSOFT’S INVOLVEMENT .....13  
VII. INTERROGATING THE OPENAI LANGUAGE MODELS USING CHATGPT .....16  
VIII. CLASS ALLEGATIONS .....17  
IX. CLAIMS FOR RELIEF .....22  
X. DEMAND FOR JUDGMENT .....36

1 Plaintiffs Catherine Denial, Ian McDowell, and Steven Schwartz, on behalf of themselves and  
2 all others similarly situated, bring this class action complaint (“Complaint”) against Defendants  
3 OpenAI, Inc.; OpenAI, L.P.; OpenAI OpCo, L.L.C.; OpenAI GP, L.L.C.; OpenAI Startup Fund I, L.P.;  
4 OpenAI Startup Fund GP I, L.L.C.; and OpenAI Startup Fund Management, LLC. (collectively,  
5 “OpenAI”) and Defendant Microsoft Corporation (“Microsoft”).

## 6 I. INTRODUCTION

7 1. In the race to dominate the emerging field of generative artificial intelligence (“GenAI”),  
8 OpenAI engaged in a systematic campaign of IP theft and data piracy. In carrying out this scheme,  
9 OpenAI engaged in unlawful conduct by copying tens of millions of copyrighted works—including  
10 articles, essays, and other written works—without the consent of the authors and content creators.  
11 OpenAI copied these works from so-called “shadow libraries”<sup>1</sup> or pirated databases that have themselves  
12 been the target of numerous legal actions brought by government enforcers for criminal copyright  
13 infringement, money laundering, and other claims. In addition to directly downloading massive amounts  
14 of pirated data, OpenAI also obtained copies of this data via peer-to-peer file-sharing networks used to  
15 facilitate data piracy. This activity violated the rights of countless authors and content creators throughout  
16 the United States and undermined foundational principles of innovation through fair competition.

17 2. OpenAI’s disregard of creators’ rights was no oversight. OpenAI sought out and torrented,  
18 for its commercial use, tens of millions of pirated copyrighted works. OpenAI copied those works without  
19 consent, credit, or compensation, and as part of this effort, pirated authors’ content through shadow  
20 libraries like Library Genesis (aka “libgen” or “LibGen”).

21 3. OpenAI’s disregard for copyright, data piracy laws, and ethical standards was not merely  
22 a case of corporate negligence. It was part of a strategy to amass a competitive advantage as fast as  
23 possible while knowingly flouting existing laws and rights that protect this country’s authors and creators.

24 4. Microsoft was and is the key business partner of OpenAI. Microsoft played a critical role  
25 in enabling and profiting from OpenAI’s unlawful activities. As a significant investor and operational  
26

---

27 <sup>1</sup> In this Complaint, “shadow libraries” refers to any online repositories or large datasets containing  
28 copyrighted material of any kind, assembled and made freely accessible online without permission, and  
in any medium, including but not limited to any copyrighted text, images, audio, video, and programming  
code.

1 partner, Microsoft provided the financial resources, cloud infrastructure, and technical support necessary  
2 for OpenAI to acquire, process, and exploit massive amounts of stolen IP. Microsoft worked closely with  
3 OpenAI in the development, testing and commercialization of OpenAI’s generative AI products.  
4 Microsoft provided OpenAI with data and environments to develop its infringing models. By integrating  
5 OpenAI’s models into its own commercial products and services, Microsoft was a key participant in the  
6 development of these products and derived substantial and direct economic benefits from their joint  
7 conduct. Microsoft acted with knowledge and directly benefited from OpenAI’s scheme. Microsoft acted  
8 jointly in the unlawful conduct at issue in this case and committed a series of overt acts and other conduct  
9 in furtherance of their scheme and common course of conduct. In addition, Microsoft and OpenAI, who  
10 are horizontal competitors in the market for training data for LLMs, formed an anti-competitive cartel,  
11 working together on an exchange of training data for LLMs, including direct copies of unlawfully  
12 acquired works and conspiring to deny sellers or potential sellers of the income that would have been  
13 received from the market for training data but for their joint conduct.

14 5. The ramifications of OpenAI’s conduct extend far beyond the immediate harm to  
15 individual copyright holders. By building its GenAI models on a foundation of stolen IP, taken without  
16 compensation, OpenAI has sought to normalize copyright infringement as the leading business strategy  
17 of the GenAI industry for obtaining text data to train their models. Microsoft’s support and integration  
18 of these unlawfully trained models into its own products magnify that impact, further foreclosing actual  
19 competition and future competition, ensuring financial benefits and preventing the entry and  
20 development of competitive market forces in the future.

21 6. Plaintiffs, representing a class of copyright owners whose works have been unlawfully  
22 acquired and exploited, seek not only to hold OpenAI and Microsoft accountable for their actions, but  
23 also to deter similar conduct in the future by other GenAI companies and bad actors who seek to exploit  
24 their works.

## 25 II. OVERVIEW

26 7. ChatGPT is a web-based software application created, maintained, and sold by OpenAI.  
27 ChatGPT is powered by AI software programs also known as *large language models* (“LLMs”). Vast  
28 quantities of data have been integral—indeed essential—to the development of these products and will

1 continue to be so in the future. There is no substitute for this data. Two of OpenAI’s most popular  
2 models are called GPT-3.5 and GPT-4. More are on the way.

3 8. LLMs, rather than being programmed in the traditional way, are “trained.” The so-called  
4 training process starts with copying massive amounts of text, often called *raw data*. That text is then  
5 processed and expressive information is extracted from it. The resulting corpus of text is called the  
6 *training dataset*. During the training process, computer engineers copy the text and program the LLM  
7 to ingest the text as part of the LLM training dataset. At the end of training, the LLM is able to mimic  
8 the expressive information found in the training dataset, thereby emitting convincingly naturalistic text  
9 output in response to user prompts. This process of copying and regurgitation is key to the basic  
10 function of LLMs. It is true of the LLMs at issue in this case.

11 9. LLM output is entirely and uniquely reliant on the material in its training dataset. In  
12 other words, every time it assembles a text output, the LLM relies on the entirety of information it  
13 extracted from its training dataset. The fact it does so is crucial to its operation.

14 10. Plaintiffs and Class members are authors of text materials, including articles, essays, and  
15 other written works, which OpenAI copied and used to train its LLMs.<sup>2</sup> Plaintiffs and Class members  
16 hold copyrights in these published works. Plaintiffs and Class members did not consent to the copying  
17 or use of their works as training data by OpenAI for its LLMs. OpenAI did not obtain permission or  
18 compensation to Plaintiffs for doing so. Instead, OpenAI copied, commercially exploited and took  
19 without compensation these valuable copyrighted materials without permission and, at times, through  
20 illegal torrenting that violates copyright and data privacy laws.

21 Defendants, individually and collectively, through the use of OpenAI’s LLMs and ChatGPT,  
22 benefit commercially and profit significantly from their use of Plaintiffs’ and Class members’  
23 copyrighted works.

---

24  
25  
26  
27 <sup>2</sup> OpenAI’s LLMs include any models in development or released commercially, even if not to public  
28 sources, and irrespective of whether those models underlie ChatGPT. For purposes of this Complaint,  
OpenAI LLMs includes all products derived by OpenAI or Microsoft from OpenAI’s LLMs.

1 **III. JURISDICTION AND VENUE**

2 11. This Court has subject-matter jurisdiction under 28 U.S.C. § 1331, including because  
3 this case arises under the Copyright Act (12 U.S.C. § 101, *et seq.*).

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

13 13. Pursuant to Civil Local Rule 3-2(c), assignment of this case to the San Francisco  
14 Division is proper because this case pertains to intellectual-property rights, which is a district-wide case  
15 category under General Order No. 44, and therefore venue is proper in any courthouse in this District.

16 **IV. PARTIES**

17 **A. Plaintiffs**

18 14. Plaintiff Catherine Denial is a writer who lives in Illinois and owns a registered  
19 copyright in multiple works, including *A proper light before the country: the shifting politics of gender*  
20 *and kinship among the Dakota, Ojibwe and non-native communities of the Upper Midwest, 1825-*  
21 *1845.*<sup>3</sup>

22 15. Plaintiff Ian McDowell is a writer who lives in North Carolina and owns copyrights in  
23 multiple works, including *Wilmington Massacre was Confederacy’s Revenge.*

24 16. Plaintiff Steven Schwartz is a writer who lives in Arizona and owns copyrights in  
25 multiple works, including *A Comprehensive System for Item Analysis in Psychological Scale*  
26 *Construction.*

27  
28 \_\_\_\_\_  
<sup>3</sup> Registration No. TX0006474253.

1 17. A non-exhaustive list of copyrights owned by Plaintiffs is shown in Exhibit A. Together,  
2 and for the purposes of this Complaint, these works are collectively referred to as the **Selected**  
3 **Infringed Works.**

4 **B. Defendants**

5 18. Defendant OpenAI, Inc. is a Delaware nonprofit corporation with its principal place of  
6 business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Inc. owns and controls the  
7 other OpenAI entities.

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

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

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

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

27 23. Defendant OpenAI Startup Fund GP I, L.L.C. (“OpenAI Startup Fund GP I”) is a  
28 Delaware limited liability company with its principal place of business located at 3180 18th Street, San

1 Francisco, CA 94110. OpenAI Startup Fund GP I is the general partner of OpenAI Startup Fund I.  
2 OpenAI Startup Fund GP I is a party to the unlawful conduct alleged herein. OpenAI Startup Fund GP I  
3 manages and operates the day-to-day business and affairs of OpenAI Startup Fund I.

4 24. Defendant OpenAI Startup Fund Management, LLC (“OpenAI Startup Fund  
5 Management”) is a Delaware limited liability company with its principal place of business located at  
6 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund Management is a party to the  
7 unlawful conduct alleged herein. OpenAI Startup Fund Management was aware of the unlawful  
8 conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

9 25. Defendant Microsoft Corporation is a Washington corporation with its principal place of  
10 business located at One Microsoft Way, Redmond, Washington 98052. It also maintains multiple  
11 offices and facilities, key employees, and a sizable customer population within this District, and it  
12 conducts business in this District.

### 13 **C. Agents and Co-Conspirators**

14 26. The unlawful acts alleged against Defendants were authorized, ordered, or performed by  
15 Defendants’ respective officers, agents, employees, representatives, or shareholders while actively  
16 engaged in the management, direction, or control of Defendants’ businesses or affairs. Defendants’  
17 agents operated under the explicit and apparent authority of their principals. Each Defendant, and its  
18 subsidiaries, affiliates, and agents, operated as a single unified entity.

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

## 23 **V. FACTUAL ALLEGATIONS**

### 24 **A. Background on OpenAI’s LLMs**

25 28. OpenAI creates, markets and sells artificial intelligence (“AI”) software products.  
26 Generally, AI software is designed to attempt to algorithmically simulate human reasoning or inference,  
27 often using statistical methods. AI models do not think or reason like humans. AI models mimic certain  
28 human interactions, including, for example, by providing answers to questions or user prompts.

1           29.     Certain AI products created and sold by OpenAI are known as *large language models*,  
2 or LLMs for short. An LLM is AI software designed to parse and emit natural-sounding text generally  
3 in response to user inquiries or prompts. Though an LLM is a software program, written by computer  
4 scientists or engineers, it is not created in the way most software programs are—that is, by human  
5 software engineers writing code. Rather, LLMs rely on training by copying massive amounts of text  
6 data from various sources and feeding these copies into a computer model at various stages of the LLM  
7 process.

8           30.     The training of an LLM begins with the collection of *raw data*. Raw data includes  
9 textual material collected from various sources—some legal (e.g., Project Gutenberg, an online  
10 repository of out-of-copyright books)—and some not (e.g., notorious shadow libraries or pirated  
11 material like LibGen). Once gathered, raw data is processed—for instance, processing can include  
12 removing low-quality raw data and organizing the dataset to make training easier. The resulting  
13 processed data comprises the *training dataset* that is fed to the LLM.

14           31.     During training, the LLM copies each piece of text in the training dataset and extracts  
15 expressive information from it. The LLM progressively adjusts its output to more closely resemble the  
16 sequences of words copied from the training dataset. Once the LLM has copied and ingested all this  
17 text, it is frequently able to emit convincing simulations of natural written language as it appears in the  
18 training dataset.

19           32.     Much of the raw data OpenAI acquired and uses in its training datasets comes from  
20 copyrighted material—encompassing a range of text data such as articles, essays, and other written  
21 works authored by Plaintiffs and other copyright holders—that were copied by OpenAI without  
22 consent, credit, or compensation, including through illegal torrenting from shadow libraries like  
23 LibGen or by crawling and scraping the internet with little to no regard for the copyright status of the  
24 scraped materials or any terms and conditions proscribing such scraping. OpenAI and Microsoft could  
25 have obtained this material legally, in compliance with copyright and other laws, but chose not to do so.

26           33.     Authors, including Plaintiffs, typically publish their works with certain copyright  
27 management information, or “CMI.” This information generally includes the title of the work, the  
28 ISBN number or copyright number, the author’s name, the copyright holder’s name, and terms and

1 conditions of use. This information is usually displayed prominently in the introductory or  
2 bibliographic sections of published materials, including articles, essays, and other written works.

3 34. OpenAI made a series of LLMs, including but not limited to GPT-1 (released June  
4 2018), GPT-2 (February 2019), GPT-3 (May 2020), GPT-3.5 (March 2022), GPT-4 (March 2023) and  
5 other variations still in development and set to be released. “GPT” is an abbreviation for “generative  
6 pre-trained transformer,” where *pre-trained* refers to the use of text data for training, *generative* refers  
7 to the model’s ability to emit text, and *transformer* refers to the underlying training algorithm. OpenAI  
8 offers certain language models in variant forms: for instance, the GPT-4 family of models includes  
9 publicly accessible variants called ‘gpt-4-0125-preview,’ ‘gpt-4-turbo-preview,’ and ‘gpt-4-32k;’ the  
10 GPT-3.5 Turbo family of models includes publicly accessible variants called ‘gpt-3.5-turbo-0125,’ ‘gpt-  
11 3.5-turbo-1106,’ and ‘gpt-3.5-turbo-instruct.’ Starting in December 2024, OpenAI also began releasing  
12 a series of “reasoning” LLMs (LLMs designed to accomplish more complex reasoning tasks like  
13 solving puzzles or riddles): o1, o1-mini, o3, and o3-mini. There are other models as well  
14 (<https://platform.openai.com/docs/models>), and OpenAI continues to develop more: In an interview  
15 with the Financial Times in November 2023, OpenAI CEO Sam Altman confirmed OpenAI was  
16 developing GPT-5. More than a year later, in a February 12, 2025 post on the social media platform X,  
17 Altman confirmed GPT-5 is still under development and said OpenAI will first release GPT-4.5.

18 35. While some of OpenAI’s LLMs and GPT language-model variants are publicly available  
19 and free to download, others require paid monthly or annual subscriptions. OpenAI has also made other  
20 language-model variants that are in commercial use and integrated into products manufactured and sold  
21 by others.

22 36. OpenAI may use many kinds of materials to train its AI systems and models. But  
23 copyrighted text data has always been a key ingredient used by OpenAI and Microsoft in training  
24 datasets for its LLMs.

25 37. In addition to the data necessary to train the LLMs, another key input is computing  
26 power. LLMs require large, fast, sophisticated computing. It has been reported that the training dataset  
27 for GPT-4 contained over 1.4 trillion tokens. Training on this data, which sometimes occurs over one or  
28 more epochs, can take days or weeks.

1 38. There is already a substantial market for AI training data with many willing buyers and  
2 sellers. The market is valued by some analysts at approximately 2.92 billion USD in 2024 and  
3 projected to exceed 17 billion USD by 2032. There is also a market for LLM training data, which  
4 includes copyrighted literary works such as fiction and non-fiction. Recognizing the economic value  
5 copyrighted works have as training data, GenAI companies have negotiated and entered into licensing  
6 agreements to use copyrighted as training data.

7 39. OpenAI and Microsoft are major players in the relevant market for LLM training data.  
8 They are also horizontal competitors in the market for LLM training data, whether it be for registered  
9 copyrighted works, or for unregistered textual works. OpenAI has entered into deals with a variety of  
10 organizations such as Axel Springer, the Financial Times, Reddit and the Associated Press to license  
11 their content as training data for its LLMs. Microsoft has also entered into licensing deals with  
12 organizations for licensing training data for LLMs, including a November 2024 deal with book  
13 publisher HarperCollins to use nonfiction works (and almost certainly copyrighted) as training data.

14 40. OpenAI and Microsoft recognize the value of the material used to train its LLMs,  
15 whether it is copyrighted works or unregistered textual material. OpenAI and Microsoft recognize that  
16 the textual material is protected by copyright and other laws that protect authors and prohibit taking of  
17 the textual material without permission or compensation. OpenAI and Microsoft knowingly and  
18 willfully violated those laws. OpenAI and Microsoft knowingly and willfully made the crass business  
19 decision to take what they could.

20 **B. OpenAI targets and steals copyrighted works**

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

28

1           42. BookCorpus, however, is an illicit dataset of pirated books. It was assembled in 2015 by  
2 a team of AI researchers for the purpose of training language models. They copied the books from the  
3 website [www.smashwords.com](http://www.smashwords.com), which makes unpublished novels available online at no cost. Those  
4 novels are largely under copyright and were copied into the BookCorpus dataset without consent,  
5 credit, or compensation to the authors.

6           43. Despite these known issues, OpenAI proceeded to copy and use BookCorpus for  
7 training its LLMs, including GPT-1. Their decision underscores a pattern of negligence and disregard  
8 for the legal and ethical standards governing the use of copyrighted materials.

9           44. OpenAI also accessed and copied vast quantities of copyrighted works, including  
10 Plaintiffs' works, through various other illegal sources, including from notorious shadow libraries such  
11 as LibGen. Sometimes, OpenAI did so by torrenting and seeding these pirated works—in other words,  
12 downloading and sharing Plaintiffs' and others' copyrighted works using peer-to-peer networks.

13           45. In the July 2020 paper introducing GPT-3 (called “Language Models are Few-Shot  
14 Learners”), OpenAI disclosed that 15% of its enormous GPT-3 training dataset came from “two  
15 internet-based books corpora,” which OpenAI pretextually referred to as “Books1” and “Books2,”  
16 concealing the names used internally by OpenAI employees when referring to these datasets: Libgen1  
17 and Libgen 2 (collectively, “LibGen Datasets”). At the time, the true source and provenance of  
18 “Books1” and “Books2” was a mystery, which OpenAI knowingly concealed.

19           46. Tellingly, OpenAI never publicly revealed which copyrighted books and other works are  
20 part of the LibGen Datasets—though there are some clues. First, OpenAI admitted these are “internet-  
21 based books corpora.” Second, the LibGen Datasets are apparently much larger than BookCorpus.  
22 Microsoft was aware that the material described had in fact been obtained from pirate websites or other  
23 illicit sources.

24           47. The only “internet-based books corpora” that have ever made that quantity of material  
25 available are shadow libraries like LibGen, Z-Library (aka B-ok), Sci-Hub, Internet Archive, and  
26 Bibliotik. These datasets are large collections of pirated materials stolen from authors around the world.  
27 *See Cengage Learning, Inc. v. Library Genesis*, Case No. 23-cv-08136 (S.D.N.Y. Sep. 24, 2024), Dkt.  
28 36 (permanently enjoining LibGen due to copyright infringement); *Hachette Book Group, Inc. v.*

1 *Internet Archive*, Case No. 20-cv-04160-JGK-OTW, (S.D.N.Y. Aug. 11, 2023), Dkt. 213 (permanently  
2 enjoining Internet Archive due to copyright infringement). OpenAI accessed these pirated databases  
3 and illegally downloaded and torrented mass quantities of copyrighted works.

4 48. After accessing and copying these stolen works, OpenAI compiled them into training  
5 datasets.

6 49. On information and belief, one or more of the Selected Infringed Works for each Plaintiff  
7 are found in OpenAI's datasets.

8 **C. OpenAI accessed and copied vast amounts of copyrighted works using peer-to-peer  
9 file sharing**

10 50. OpenAI's use of LibGen demonstrates that OpenAI knowingly and intentionally torrented  
11 large volumes of digital files containing pirated copyrighted works, including Plaintiffs' works.

12 51. OpenAI's reliance on torrenting is especially alarming because obtaining data through  
13 peer-to-peer sharing generally involves not just copying and hosting pirated data, but uploading,  
14 distributing or "seeding" pirated data. In other words, to acquire a torrented file, a user must typically  
15 participate in a data exchange: data is downloaded from fellow pirates while simultaneously uploaded  
16 to fellow pirates. Thus, it is plausible that OpenAI was not only downloading and copying massive  
17 amounts of pirated copyrighted works but also distributing them to other IP pirates in the swarm.

18 52. Microsoft knew, or should have known, that OpenAI had obtained pirated copyrighted  
19 works by torrenting.

20 **D. OpenAI attempts to conceal its use of torrented copyrighted data**

21 53. OpenAI tried to hide its piracy in at least two ways.

22 54. *First*, at the individual file level, OpenAI wanted to conceal and obscure its reliance on  
23 copyrighted data by stripping copyright-identifying information from the files it stole.

24 55. *Second*, and more broadly, OpenAI sought to obscure the origins of the pirated data it  
25 accessed and copied for use with its LLMs. For example, OpenAI coined the term "Books1" and  
26 "Books2," concealing the true names for these datasets: Libgen1 and Libgen 2 (collectively, "LibGen  
27 Datasets").  
28

1 56. OpenAI’s concealment of the data it acquired and processed into training datasets for its  
2 LLMs continued for years. In March 2023, OpenAI’s paper introducing GPT-4 contained no  
3 information about its dataset at all, claiming that “[g]iven both the competitive landscape and the safety  
4 implications of large-scale models like GPT-4, this report contains no further details about . . . dataset  
5 construction.”

6 57. Microsoft knew, or should have known, that OpenAI concealed its torrenting of pirated  
7 copyrighted data and provided incorrect pretextual explanations. Despite this knowledge or  
8 information, Microsoft did not disclose these facts.

9 **E. OpenAI steals additional copyrighted material by crawling and scraping the internet**

10 58. OpenAI’s LLM training datasets have included a wide array of other copyrighted  
11 materials, such as articles, essays, and other written works, taken without permission. OpenAI has used  
12 data repositories such as Common Crawl, a publicly available dataset that scrapes vast amounts of  
13 internet content, including copyrighted material from websites, blogs, and news articles. This extensive  
14 use of Common Crawl derived data underscores the breadth of copyrighted materials taken by OpenAI  
15 and ingested by its models, which extends far beyond books to encompass a diverse range of written  
16 works.

17 59. In contrast with its circumspection about shadow libraries, OpenAI *has* publicly  
18 admitted to using Common Crawl to develop its LLMs. In a paper authored by several AI researchers,  
19 including OpenAI engineers who worked directly on GPT-3, the downloading and use of Common  
20 Crawl is discussed openly. This paper, *Language Models are Few-Shot Learners* by Tom B. Brown et  
21 al.,<sup>4</sup> admits that “The CommonCrawl data was downloaded from 41 shards of monthly CommonCrawl  
22 covering 2016 to 2019, constituting 45TB of compressed plaintext . . .” This extensive dataset includes  
23 a vast array of copyrighted text from websites, blogs, and news articles, highlighting the breadth of  
24 sources used to train OpenAI’s models.

25 60. Common Crawl publishes insights into each of the “crawls” they conduct, including the  
26 top domains included in each dataset. The insights published by Common Crawl on the May through  
27

28 \_\_\_\_\_  
<sup>4</sup> <https://arxiv.org/pdf/2005.14165>

1 June 2018 crawl reveal that the domains WordPress.com and BlogSpot.com were both in the top 20  
2 domains crawled for data.<sup>5</sup> Those platforms host millions of blogs and articles, many of which are  
3 copyrighted. The inclusion of such domains in the Common Crawl dataset underscores OpenAI's  
4 extensive copying and use of diverse copyrighted materials, including articles, essays, and other written  
5 works with its LLMs.

6 61. Because Common Crawl copies essentially the entire internet, on information and belief,  
7 the Selected Infringed Works (or parts of them) can be found in OpenAI's common crawl datasets and  
8 similar datasets that are the product of scraping and crawling.

9 **F. OpenAI and Microsoft knowingly profit from stealing copyrighted material**

10 62. OpenAI's practices in acquiring text data for training its LLMs have come under  
11 significant scrutiny, particularly regarding its use of peer-to-peer file-sharing networks to acquire  
12 massive quantities of copyrighted material from shadow libraries such as LibGen. Downloading,  
13 including by torrenting, pirated IP is not only unlawful but also negates any attempt to claim fair use.

14 63. At every turn, when faced with what it saw as a choice to respect copyright law and  
15 intellectual property rights or gain competitive advantage, OpenAI knowingly chose the latter.

16 **VI. MICROSOFT'S INVOLVEMENT**

17 64. Defendant Microsoft has played a significant role in the development and operation of  
18 OpenAI, both financially and operationally. Microsoft has invested billions of dollars into OpenAI.  
19 This financial backing not only provided OpenAI with the resources necessary to develop its AI  
20 systems, but has also given Microsoft knowledge and awareness of OpenAI's conduct. It has also  
21 provided Microsoft substantial influence over, and even the ability to control, OpenAI's operations and  
22 strategic decisions. Microsoft has sought, obtained and maintained such influence because Microsoft  
23 attempted and failed to design its own LLM and other AI products. In order to support its business  
24 strategy, Microsoft provided OpenAI with infrastructure such as access to its cloud network and  
25 computing power in order for OpenAI to develop its infringing technology. Without Microsoft, OpenAI  
26 would not have been able to have done so.

27  
28 

---

<sup>5</sup> <https://commoncrawl.github.io/cc-webgraph-statistics/>

1           65. In addition, incorporation of OpenAI’s GenAI technology into Microsoft’s commercial  
2 products and services, such as Azure, Microsoft Office, and other enterprise solutions was a key  
3 component of the partnership. Microsoft and OpenAI were also commercializing other products,  
4 including Microsoft’s Bing search engine. Microsoft’s incorporation of OpenAI’s LLM technology was  
5 a central pillar of both companies’ commercialization of their LLM technology. This integration has  
6 allowed Microsoft to benefit directly from the AI systems developed by OpenAI, including the use of  
7 copyrighted material obtained unlawfully in developing and training its LLMs. This required  
8 substantial integration of the software engineers and computer scientists of the two companies. Indeed,  
9 particular teams of officers and other employees of the companies were formed to work together.  
10 Among other things, these joint endeavors provided the knowledge and information, shared by the two  
11 companies, of the unlawful conduct at issue in this case. By incorporating OpenAI’s models into its  
12 products, Microsoft has effectively directly endorsed and profited from OpenAI’s activities.

13           66. Given Microsoft’s deep involvement in OpenAI’s operations, Microsoft was aware of,  
14 and upon information and belief, approved, OpenAI’s piracy and other unlawful data acquisition  
15 practices. Public statements and internal communications suggest that Microsoft had access to  
16 information about the sources of OpenAI’s training data.

17           67. Microsoft was not just aware of OpenAI’s approach; Microsoft encouraged it and  
18 profited from it because Microsoft received OpenAI’s training data. On information and belief,  
19 Microsoft received that data as part of a trade in which Microsoft provided OpenAI with Bing crawl  
20 data and OpenAI provided Microsoft with its training data. On information and belief, OpenAI’s side of  
21 the trade included LibGen.

22           68. Incredibly, in a joint press release in July 2019, Microsoft and OpenAI stated: “We are  
23 dedicated to ensuring that our AI technologies are developed and used in a manner that is ethical and  
24 respects the rights of all individuals.” (Microsoft-OpenAI Press Release, July 2019.) Hardly. On  
25 information and belief, at this very time, OpenAI, with Microsoft’s knowledge, was illegally torrenting  
26 massive amounts of copyrighted works from shadow libraries such as LibGen, in addition to web  
27 crawling and scraping additional copyrighted works, in violation of the rights of copyright owners like  
28

1 Plaintiffs. Microsoft condoned and supported OpenAI’s mass IP piracy, fully aware of its legal  
2 implications.

3 69. Worse, by blessing—and profiting from—OpenAI’s data theft, Microsoft and OpenAI  
4 artificially limited the market for training data. In a well-functioning market, OpenAI and Microsoft  
5 would compete to buy and license high-quality training data from copyright holders. But OpenAI and  
6 Microsoft suppressed that market by sharing stolen data—essentially agreement to a price of zero.

7 70. The companies independently and together recognized that one of their shared common  
8 purposes was to limit the market for AI training data. As part of their agreement to work together—  
9 including OpenAI’s agreement to provide Microsoft stolen copyrighted works—Microsoft agreed to  
10 reduce, prevent and foreclose competition in the AI training space. Had the two companies not  
11 embarked on their common course of conduct, the market would have grown and developed, and  
12 owners of copyrights could have, and would have, been able to obtain fair value for their work in a  
13 marketplace where prices would have been set by willing buyers and sellers.

14 71. Microsoft eventually began to assist OpenAI in its data acquisition efforts. Microsoft  
15 ultimately obtained and shared with OpenAI training data derived, on information and belief, from  
16 scraping and crawling the public internet.

17 72. Microsoft’s stated goal was to acquire AI training data legally through business  
18 partnerships.

19 73. Yet, on information and belief, Microsoft also sought to acquire and use vast quantities  
20 of pirated works from LibGen for training purposes.

21 74. Microsoft’s Azure cloud platform has also been a critical infrastructure component for  
22 OpenAI, providing the computing power necessary to train OpenAI’s LLMs. By providing this  
23 infrastructure, Microsoft facilitated the processing and storage of the vast amounts of data acquired and  
24 used by OpenAI, including the vast quantities of copyrighted works that OpenAI illegally torrented  
25 from shadow libraries of pirated material. Microsoft’s role in providing the technological backbone for  
26 OpenAI’s operations implicates it in the unlawful data-acquisition scheme employed by OpenAI. Not  
27 only is Microsoft providing critical infrastructure to store pirated works, but it intends to allow OpenAI  
28

1 the use of the Azure cloud platform for the next decade without any indication that it will stop storing  
2 these pirated works.

3 75. Microsoft has made several other public statements which provide pretextual and  
4 inaccurate versions of OpenAI's practices of acquiring AI training data. For example, Microsoft CEO  
5 Satya Nadella wrote, "We are committed to the highest standards of data ethics and transparency in all  
6 our AI endeavors." (Microsoft Blog, January 2021) (a line that OpenAI CEO Sam Altman echoed when  
7 he told the Financial Times in November 2023, "We are committed to using only publicly available  
8 data and data we have the right to use"). Such statements were knowingly deceptive and misleading  
9 given the evidence that OpenAI, under Microsoft's substantial influence and support, purposely,  
10 repeatedly, and illegally obtained and used pirated copyrighted material to train its AI models.

11 76. Additionally, Microsoft claimed that its partnership with OpenAI was built on a  
12 foundation of "trust and integrity" (Microsoft Annual Report, 2022), despite knowing full well the  
13 illegal data-acquisition methods employed by OpenAI. Microsoft internal personnel knew, or should  
14 have known, that this was at odds with the truth, and Microsoft's own internal strategy and designs. By  
15 providing the financial resources, technological infrastructure, and strategic support necessary for  
16 OpenAI to develop its models, Microsoft has contributed materially to the unauthorized accessing and  
17 infringement of Plaintiffs' copyrighted works and conspired alongside OpenAI to violate Plaintiffs'  
18 rights.

19 77. Put another way, OpenAI and Microsoft shared a common plan and purpose: to access  
20 Plaintiffs' copyrighted works without authorization, infringe on Plaintiffs' copyrights, and conceal both  
21 the origins of the data pirated to train OpenAI's and Microsoft's AI models and the means with which it  
22 was acquired. Microsoft knew, or should have known, the reasonable and foreseeable results of this  
23 conduct.

## 24 **VII. INTERROGATING THE OPENAI LANGUAGE MODELS USING CHATGPT**

25 78. ChatGPT is an LLM created and sold by OpenAI. As its name suggests, ChatGPT is  
26 designed to offer a conversational style of interaction with a user. OpenAI offers ChatGPT through a  
27 web interface to users for free and via paid subscriptions.  
28



1 **All persons or entities domiciled in the United States that own a United States**  
2 **copyright in any textual work, where the work is registered with the United**  
3 **States Copyright Office, but are not assigned one or more International Standard**  
4 **Books Number(s) (ISBN) or Amazon Standard Identification Number(s) (ASIN).**

5 This Class definition excludes:

- 6 a. any of the Defendants named herein;
- 7 b. any of the Defendants' co-conspirators;
- 8 c. any of Defendants' parent companies, subsidiaries, and affiliates;
- 9 d. any of Defendants' officers, directors, management, employees, subsidiaries,  
10 affiliates, or agents;
- 11 e. all governmental entities; and
- 12 f. the judges and chambers staff in this case, including on appeal, as well as any  
13 members of their immediate families. This exclusion applies regardless of the  
14 type of copyrighted text material involved.

15 85. 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 ("Unregistered Copyright Holders Class"):

18 **All persons or entities domiciled in the United States that own a United States**  
19 **copyright in any text data, who have not registered their works with the United**  
20 **States Copyright Office, including but not limited to books, articles, essays, and**  
21 **other written works, that was accessed, copied, or used by OpenAI during the**  
22 **Class Period.**

23 This Class definition excludes:

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

1 members of their immediate families. This exclusion applies regardless of the type of  
2 copyrighted text material involved.

3 86. **Numerosity.** Plaintiffs do not know the exact number of members in each Class. This  
4 information is in the exclusive control of Defendants. On information and belief, there are at least  
5 hundreds of thousands of members in the Classes geographically dispersed throughout the United  
6 States, encompassing owners of a wide range of copyrighted text materials, including articles, essays,  
7 and other written works. Therefore, joinder of all members of each Class in the prosecution of this  
8 action is impracticable.

9 87. **Typicality.** Plaintiffs' claims are typical of the claims of other members of each Class  
10 because Plaintiffs and all members of each Class were damaged by the same wrongful conduct of  
11 Defendants as alleged herein, including the unauthorized access and CMI stripping. The relief sought  
12 herein also is common to all members of each Class.

13 88. **Adequacy.** Plaintiffs will fairly and adequately represent the interests of the members of  
14 the respective Classes because the Plaintiffs have experienced the same harms as the members of the  
15 Class, including the unauthorized access and CMI stripping, and Plaintiffs have no conflicts with any  
16 other members of the Class. Furthermore, Plaintiffs retained and are represented by sophisticated and  
17 competent counsel who are experienced in prosecuting federal and state class actions, as well as other  
18 complex litigation.

19 89. **Commonality and predominance.** Numerous questions of law or fact common to each  
20 Class arise from Defendants' conduct:

- 21 a. whether Defendants' conduct alleged herein, including but not limited to the  
22 unlawful use of peer-to-peer file-sharing networks to access and copy pirated  
23 copyrighted works and the misrepresentation of data sources used to train  
24 OpenAI's models, constitutes Unfair Competition under California Business and  
25 Professions Code § 17200 *et seq.*;
- 26 b. whether this Court should enjoin Defendants from engaging in the unlawful  
27 conduct alleged herein, including the unauthorized access, copying, and use of  
28 Plaintiffs' copyrighted material, and what the scope of that injunction would be,

- 1 including but not limited to whether OpenAI and Microsoft should be allowed to  
2 continue offering their suite of products trained on Plaintiffs' works unlawfully;
- 3 c. whether Defendants' actions in copying mass quantities of text material from the  
4 internet, including but not limited to Plaintiff Catherine Denial's work, *A proper*  
5 *light before the country: the shifting politics of gender and kinship among the*  
6 *Dakota, Ojibwe and non-native communities of the Upper Midwest, 1825-1845,*  
7 without Plaintiffs' permission, constitute direct copyright infringement under 17  
8 U.S.C. § 501;
- 9 d. whether Microsoft, by providing financial resources, technical infrastructure, and  
10 strategic support to OpenAI, had the right and ability to supervise and control  
11 OpenAI's infringing activity and failed to exercise such supervision and control;
- 12 e. whether Defendants' conduct, including but not limited to, the unauthorized use  
13 of peer-to-peer file-sharing networks to access and copy pirated copyrighted  
14 works and the misrepresentation of data sources used to train OpenAI's models,  
15 constitutes Unfair Competition under California Business and Professions Code  
16 § 17200 *et seq.*;
- 17 f. whether OpenAI's unauthorized access and use of Plaintiffs' copyrighted  
18 Infringed Works, including but not limited to by torrenting digital copies from  
19 shadow libraries such as LibGen, constitute violations of the California  
20 Comprehensive Computer Data Access and Fraud Act (CDAFA), Cal. Penal  
21 Code § 502;
- 22 g. whether OpenAI circumvented technological measures that control access to  
23 Plaintiffs' copyrighted works, including by torrenting these and other  
24 copyrighted materials from shadow libraries such as LibGen, in violation of the  
25 Digital Millennium Copyright Act (DMCA), 17 U.S.C. § 1201, and whether  
26 OpenAI's removal of copyright management information (CMI) from  
27 copyrighted works, including but not limited to the Selected Infringed Works,  
28 constitutes a violation of the DMCA, 17 U.S.C. § 1201(b)(1);

- 1 h. whether OpenAI’s unauthorized acquisition and use of Plaintiffs’ copyrighted  
2 works, including but not limited to, by torrenting them from shadow libraries  
3 such as LibGen constitute conversion under California law;
- 4 i. whether OpenAI has been unjustly enriched by its unauthorized access, copying,  
5 and use of Plaintiffs’ copyrighted material to train its GenAI models, deriving  
6 significant commercial benefits and profits from this use, and whether Microsoft  
7 directly benefited from this unjust enrichment by integrating OpenAI’s models  
8 into its own products and services, leveraging Plaintiffs’ unlawfully obtained  
9 intellectual property to enhance its offerings and increase its profits;
- 10 k. whether OpenAI’s unauthorized access and use of data from websites that  
11 prohibit such activities in their terms and conditions constitute violations of the  
12 Computer Fraud and Abuse Act (CFAA), 18 U.S.C. § 1030, and whether  
13 Microsoft contributed to these actions by providing infrastructure, financial  
14 support, and resources necessary for OpenAI to engage in these unlawful  
15 activities; and
- 16 l. whether Defendants unlawfully acquired copyrighted material from the internet,  
17 including but not limited to by torrenting works from shadow libraries and  
18 violating websites’ terms and conditions, and used the material to develop GenAI  
19 models and products in order to generate profits, in violation of California Penal  
20 Code § 496(a), (c).

21 These and other questions of law and fact are common to each Class and predominate over any  
22 questions affecting the Class members individually, particularly given the widespread and systematic  
23 nature of Defendants’ unauthorized access, copying, and use of a vast amount of copyrighted text data  
24 and works.

25 90. **Other class considerations.** Defendants acted on grounds generally applicable to the  
26 Class. This class action is superior to alternatives, if any, for the fair and efficient adjudication of this  
27 controversy. Prosecuting the claims pleaded herein as a class action will eliminate the possibility of  
28 repetitive litigation and inconsistent results. There will be no material difficulty in the management of

1 this case as a class action. Furthermore, final injunctive relief is appropriate with respect to the Class as  
2 a whole, given the systematic and widespread nature of Defendants’ unauthorized and unlawful access,  
3 copying, and use of copyrighted text data and works.

4 91. The prosecution of separate actions by individual Class members would create the risk  
5 of inconsistent or varying adjudications, establishing incompatible standards of conduct for Defendants,  
6 particularly given the widespread and systematic nature of Defendants’ unauthorized and unlawful  
7 access, copying, and use of a vast amount of copyrighted text data and works.

8 **IX. CLAIMS FOR RELIEF**

9 **COUNT 1**

10 **Direct Copyright Infringement**

11 **17 U.S.C. § 501**

12 **(Against OpenAI and Microsoft on Behalf of the Non-Book Infringement Class)**

13 92. Plaintiffs incorporate by reference the preceding factual allegations.

14 93. Plaintiffs hold the exclusive rights to works, including but not limited to the Selected  
15 Infringing Works, under 17 U.S.C. § 106.

16 94. Plaintiffs never authorized OpenAI or Microsoft to make copies of these texts, including  
17 but not limited to the Selected Infringed Works, or any portion thereof, to make derivative works, to  
18 publicly display copies (or derivative works), or to distribute copies (or derivative works). All those  
19 rights belong exclusively to Plaintiffs under copyright law.

20 95. On information and belief, in connection with training its LLMs, OpenAI and Microsoft  
21 copied mass quantities of text material, including but not limited to the Plaintiff Catherine Denial’s  
22 work, *A proper light before the country: the shifting politics of gender and kinship among the Dakota,*  
23 *Ojibwe and non-native communities of the Upper Midwest, 1825-1845*, in digital formats, including by  
24 torrenting them from one or more shadow libraries or pirate websites, such as LibGen.

25 96. To the extent not already specified, Plaintiffs incorporate by reference Exhibit A, which  
26 identifies by title, author, and (where applicable), the registration number for the copyrighted work at  
27 issue. Plaintiffs allege, on information and belief, that each of these works, or substantial portions  
28 thereof, were included in the datasets used by Defendants to train their large language models, as

1 evidenced by the ability of those models to generate verbatim or near-verbatim excerpts from said  
2 works upon prompt. Plaintiff Catherine Denial further alleges that Defendants' acts of copying, storing,  
3 and using these works in the course of training and deploying their commercial AI products constitute  
4 unauthorized reproductions in violation of 17 U.S.C. § 106.

5 97. OpenAI and Microsoft made additional copies of and/or from texts including, but not  
6 limited to, the Selected Infringed Works during its LLM training process without Plaintiffs' permission.

7 98. Licensing copyrighted material to train AI models is plainly feasible. It already happens.  
8 Indeed, OpenAI itself has licensed copyrighted material for training its LLMs. For instance, OpenAI  
9 reached agreements with the Associated Press and Axel Springer to license text data and material for its  
10 LLM training. OpenAI has reportedly been in negotiations with other publishers as well. Microsoft has  
11 also sought to negotiate and obtain licenses for text works as training data for training LLMs.

12 99. Microsoft played a pivotal role in facilitating OpenAI's infringing activities by  
13 providing the financial resources, technical infrastructure, and strategic support necessary for OpenAI  
14 to develop and expand its AI systems. This includes the use of Microsoft Azure, which powered the  
15 large-scale training of OpenAI's models using Plaintiffs' (and others') copyrighted material without  
16 authorization.

17 100. Microsoft and OpenAI also engaged in exchanges of training data. OpenAI received  
18 training data from Microsoft based on various web scrapes of the internet. Microsoft also received  
19 training data, including direct copies of the copyrighted works, directly from OpenAI for its own use.

20 101. On information and belief, OpenAI and Microsoft's infringing conduct was and  
21 continues to be willful, continuing to infringe on Plaintiff Catherine Denial and Plaintiffs and members  
22 of the Non-Book Infringement Class' exclusive rights knowing they were profiting from widescale  
23 copyright infringement.

24 102. Plaintiff Catherine Denial and members of the Non-Book Infringement Class have been  
25 injured by Defendants' acts of direct copyright infringement. Plaintiff Catherine Denial and members of  
26 the Non-Book Infringement Class are entitled to statutory damages, actual damages, restitution of  
27 profits, and/or other remedies provided by law.  
28

COUNT 2

**Vicarious Copyright Infringement**

**17 U.S.C. § 501**

**(Against Microsoft, OpenAI Inc. and OpenAI GP LLC on Behalf of the Non-Book Infringement Class)**

103. Plaintiffs incorporate by reference the preceding factual allegations.

104. As explained above in Count 1, OpenAI directly infringed copyrights owned by Plaintiff Catherine Denial and members of the Non-Book Infringement Class.

105. Microsoft directly benefitted from that infringement both because of its partnership with OpenAI and because it incorporates OpenAI’s products—which infringe Plaintiffs’ copyrights—into Microsoft’s products.

106. Defendant Microsoft, by virtue of its substantial investment, contractual arrangements, and operational integration with OpenAI, had both the legal right and practical ability to supervise and control the infringing activities of OpenAI, including but not limited to the acquisition and use of copyrighted works in training datasets. Microsoft derived a direct financial benefit from the infringement by incorporating the resulting AI models into its own commercial products and services, thereby increasing its revenues and market share. Similarly, OpenAI Inc. and OpenAI GP LLC exercised day-to-day control over the operations of OpenAI OpCo LLC and directly benefitted from the infringing activities through increased valuation, licensing revenues, and other commercial advantages.

107. Microsoft—based on its partnership with, and significant investment in, OpenAI—had the right and ability to supervise and control OpenAI’s infringing activity.

108. Essentially, instead of Microsoft entering the market and buying copyrighted works for its own training data, it received those same works from OpenAI, who had previously stolen them, and then provided them to Microsoft in exchange for data Microsoft had from its Bing search engine team’s work on data acquisition.

109. Microsoft failed to exercise its supervision and control to prevent and/or stop OpenAI’s infringement.

1 110. OpenAI Inc. and OpenAI GP LLC had the right and ability to control the direct  
2 infringement alleged in Count I because OpenAI Inc. fully controls OpenAI GP LLC, and OpenAI GP  
3 LLC fully controls OpenAI OpCo LLC, according to the corporate structure outlined above.

4 111. OpenAI Inc. and OpenAI GP LLC have a direct financial interest in the direct  
5 infringement alleged in Count I because they benefit from the profits and investments generated by  
6 OpenAI OpCo LLC's infringing activities.

7 112. OpenAI Inc. and OpenAI GP LLC failed to exercise its supervision and control to  
8 prevent and/or stop OpenAI OpCo LLC.

9 **COUNT 3**

10 **UCL — Unfair Competition**

11 **Cal. Bus. & Prof. Code §§ 17200 et seq.**

12 **(Against All Defendants on Behalf of the Non-Book Infringement Class and the Unregistered**  
13 **Class)**

14 113. Plaintiffs incorporate by reference the preceding factual allegations.

15 114. Defendants' conduct constitutes unlawful business practices under the UCL by violating  
16 the Copyright Act, the DMCA, and the CDAFA, as alleged herein. Defendants' conduct is also unfair in  
17 that it offends established public policy and is immoral, unethical, oppressive, and unscrupulous,  
18 causing substantial injury to Plaintiffs and the Class that is not outweighed by any countervailing  
19 benefits. Plaintiffs and Class members have suffered injury in fact and lost money or property as a  
20 result of Defendants' conduct, including but not limited to the deprivation of licensing revenue,  
21 diminution in the value of their copyrighted works, and loss of control over the use and dissemination  
22 of their intellectual property.

23 115. Defendants engaged in unfair business practices by, among other things, acquiring  
24 Plaintiffs' works through unlawful means, including torrenting vast amounts of pirated copyrighted  
25 works from shadow libraries; removing copyright identification information from Plaintiffs' works; and  
26 using the copies of those works to acquire additional training data from Microsoft.

27 116. Defendants also misrepresented their adherence to ethics and respect for rights with  
28 respect to their AI operations, including but not limited to their procurement and use of data.

1 117. The unfair business practices described herein violate California Business and  
2 Professions Code § 17200 *et seq.* (the “UCL”) and are unfair, unlawful, and fraudulent.

3 118. Microsoft directly contributed to these unfair business practices by providing substantial  
4 financial resources, cloud infrastructure, and strategic support to OpenAI, enabling the development of  
5 ChatGPT using unlawfully obtained copyrighted text data and material. Microsoft has further engaged  
6 in unfair practices by integrating OpenAI’s infringing models into its own commercial products and  
7 services, thereby deriving its own profits from the exploitation of Plaintiffs’ stolen copyrighted text and  
8 material.

9 119. The unfair business practices described herein violate the UCL because they are unfair,  
10 immoral, unethical, oppressive, unscrupulous, or injurious to consumers. Defendants unfairly profit  
11 from and take credit for developing a commercial product based on unattributed reproductions of those  
12 stolen writings and ideas.

13 120. The unlawful business practices described herein violate the UCL because consumers  
14 are likely to be deceived by them. Defendants knowingly and secretly acquired, copied, and trained  
15 ChatGPT using unauthorized and infringing copies of Plaintiffs’ copyrighted text. Defendants  
16 deceptively marketed their product in a manner that fails to attribute the success of their product to  
17 copyrighted material on which it is based.

18 **COUNT 4**

19 **Violation of the California Comprehensive Computer Data Access and Fraud Act (CDAFA)**

20 **Cal. Penal Code § 502**

21 **(Against Defendant OpenAI on Behalf of the Unregistered Class)**

22 121. Plaintiffs incorporate by reference the preceding factual allegations.

23 122. Defendant OpenAI, without permission, knowingly accessed and used Plaintiffs’  
24 copyrighted works by circumventing technological barriers and downloading digital copies from  
25 shadow libraries such as LibGen, in violation of Cal. Penal Code § 502(c)(1), (2), and (7). As a direct  
26 and proximate result, Plaintiffs suffered damage and loss, including but not limited to the costs incurred  
27 in investigating the unauthorized access and the diminution in value of their intellectual property.

28 123. OpenAI’s unauthorized access and use of Plaintiffs’ copyrighted works by torrenting

1 digital copies of those works from shadow libraries such as LibGen constitute violations of the  
2 California Comprehensive Computer Data Access and Fraud Act (CDAFA), Cal. Penal Code § 502.

3 124. OpenAI knowingly and without permission accessed and used data from Plaintiffs’  
4 copyrighted works to train its AI models, thereby causing harm to Plaintiffs.

5 125. OpenAI’s knowing and unauthorized access to Plaintiffs’ copyrighted works was a  
6 substantial factor in causing Plaintiffs’ harm, including, but not limited to, the loss of control over their  
7 copyrighted material and the unauthorized use of their intellectual property.

8 126. As a direct and proximate result of OpenAI’s actions, Plaintiffs have suffered damages,  
9 including, but not limited to, the loss of control over their copyrighted material and the unauthorized  
10 use of their intellectual property, as well as the amount spent to investigate or verify whether Plaintiffs’  
11 data was or was not altered, damaged, or deleted by OpenAI. Plaintiffs have engaged in and continue to  
12 engage in protracted efforts to determine how Defendants acquired their copyrighted data. On  
13 information and belief, and based on how a user typically obtains a torrent file, it is plausible that  
14 OpenAI redistributed that data through seeding and leeching, making it available to data pirates  
15 worldwide and thus furthering such piracy beyond their own downloading efforts.

16 127. Plaintiffs are entitled to compensatory damages, injunctive relief, and other equitable  
17 remedies as provided by Cal. Penal Code § 502(e).

18 **COUNT 5**

19 **Violation of the Digital Millennium Copyright Act (DMCA)**

20 **U.S.C. § 1201**

21 **(Against all Defendants on Behalf of the Non-Book Infringement Class and the Unregistered**  
22 **Class)**

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

24 129. Defendants circumvented technological measures that effectively control access to  
25 Plaintiffs’ copyrighted works, including but not limited to digital rights management systems and  
26 password protections in violation of 17 U.S.C. § 1201(a).

27 130. Defendants circumvented technological measures that control access to Plaintiffs’  
28 copyrighted works, including, but not limited to, by torrenting these and vast amounts of other

1 copyrighted material from shadow libraries such as LibGen and by ignoring robots.txt files (the  
2 filename used for implementing the Robots Exclusion Protocol, which is designed to indicate which  
3 websites crawlers are allowed to visit) and other access-related security measures and/or by using data  
4 obtained by similarly bypassing security measures.

5 131. Defendants’ conduct in bypassing these technological measures was done without  
6 authorization and for the purpose of infringing Plaintiffs’ copyrights.

7 132. Upon information and belief, based on how a user typically obtains a torrent file, it is  
8 plausible that Defendants also distributed these works, including, but not limited to, through torrenting  
9 and seeding.

10 133. As a result of these violations of the DMCA, Plaintiffs have suffered and will continue  
11 to suffer irreparable harm and are entitled to injunctive relief, statutory damages, and other remedies as  
12 provided by 17 U.S.C. § 1203.

13 **COUNT 6**

14 **CMI-Stripping: Violation of the Digital Millennium Copyright Act (DMCA)**

15 **U.S.C. § 1201(b)(1)**

16 **(Against All Defendants on Behalf of the Non-Book Infringement Class and the Unregistered**  
17 **Class)**

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

19 135. OpenAI repeatedly and intentionally removed copyright management information  
20 (“CMI”) from copyrighted works, including but not limited to the Selected Infringed Works, that  
21 OpenAI copied and used to train ChatGPT.

22 136. OpenAI source code specifically references an approach that “helps to eliminate  
23 copyright info from [the] state of [the document]” with respect to the LibGen dataset.<sup>6</sup>

24 137. Indeed, OpenAI removed CMI from Selected Infringed Works in part to enable and to  
25 facilitate infringement. Removal of CMI made it easier to use these Works as training data and because  
26 their use in training constitutes an infringement, the CMI removal “facilitated” that infringement.  
27

28 

---

<sup>6</sup>OPCO\_AG\_SRC\_CODE00000941.

1 138. OpenAI also removed CMI from texts including, but not limited to, the Selected  
2 Infringed Works contained in OpenAI’s training datasets to conceal OpenAI’s infringement of  
3 copyrighted material, including but not limited to the Selected Infringed Works, from ChatGPT users  
4 and the public.

5 139. OpenAI sought to conceal its infringement to minimize risks that ChatGPT users and the  
6 public might learn or perceive that it had engaged in mass IP piracy, copyright infringement, and other  
7 unlawful activity in developing ChatGPT. OpenAI knew that ChatGPT could generate verbatim text  
8 from copyrighted material used to train ChatGPT. Open AI also knew that ChatGPT was prone to  
9 memorizing and generating outputs of CMI unless it was removed from the copyrighted works used to  
10 train ChatGPT.

11 140. Due to, among other things, the CMI that OpenAI removed from copyrighted works and  
12 OpenAI’s knowledge that LibGen contained copyrighted articles and other textual works, OpenAI  
13 knew or had reasonable grounds to know that its removal of CMI from ChatGPT’s training data would  
14 induce, enable, facilitate, or conceal its own copyright infringement or the copyright infringement of  
15 others. Among other things, OpenAI knew or had reasonable grounds to know its removal of CMI  
16 would reduce the chances that Plaintiffs and Class members would discover OpenAI had copied texts  
17 including but not limited to the Selected Infringed Works and/or used them to train ChatGPT.

18 141. Defendants also distributed these works, including by trading training data between  
19 OpenAI and Microsoft.

20 **COUNT 7**

21 **Conversion**

22 **(Against Defendant OpenAI on Behalf of the Unregistered Class)**

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

24 143. OpenAI’s unauthorized acquisition and use of Plaintiffs’ copyrighted works by  
25 unlawfully scraping and/or crawling and/or illegally torrenting them from shadow libraries such as  
26 LibGen constitute conversion under California law.

27 144. OpenAI took Plaintiffs’ copyrighted works without permission. In doing so, OpenAI  
28 wrongfully exercised dominion and control over Plaintiffs’ property, depriving them of their rights to

1 use and control their works. OpenAI’s unauthorized acquisition and use of Plaintiffs’ copyrighted  
2 works deprived Plaintiffs of the exclusive right to control, license, and exploit their works, including  
3 the right to determine the terms and conditions of use. By unlawfully appropriating and using the works  
4 in a manner inconsistent with Plaintiffs’ rights, OpenAI exercised wrongful dominion and control over  
5 specific, identifiable digital files embodying Plaintiffs’ intellectual property, resulting in actual and  
6 substantial interference with Plaintiffs’ property interests.

7 145. As a result of OpenAI’s conversion, Plaintiffs suffered damages, including but not  
8 limited to the loss of control over their copyrighted works and the unauthorized use of their intellectual  
9 property.

10 146. Plaintiffs are entitled to compensatory damages, punitive damages, and other equitable  
11 relief as provided by California law.

## 12 **COUNT 8**

### 13 **Unjust Enrichment / Quasi-Contract**

#### 14 **(Against All Defendants on Behalf of the Unregistered Class)**

15 147. Plaintiffs incorporate by reference the preceding factual allegations.

16 148. In the alternative to Plaintiffs’ statutory and tort claims, Plaintiffs allege that Defendants  
17 have been unjustly enriched by their unauthorized acquisition, use, and exploitation of Plaintiffs’  
18 copyrighted works, from which Defendants derived substantial commercial benefits. It would be  
19 inequitable for Defendants to retain the profits and advantages obtained through such conduct without  
20 compensating Plaintiffs for the value of their intellectual property.

21 149. OpenAI has been unjustly enriched by its unauthorized access, copying, and use of  
22 Plaintiffs’ copyrighted material to train its GenAI models, deriving significant commercial benefits and  
23 profits from this use.

24 150. Microsoft directly benefited from this unjust enrichment by integrating OpenAI’s  
25 models into its own products and services, leveraging Plaintiffs’ unlawfully obtained intellectual  
26 property to enhance its offerings and increase its profits.

27 151. Defendants’ enrichment came at the expense of Plaintiffs, who have not been  
28 compensated for the acquisition and use of their copyrighted material.





COUNT 11

**Larceny/Receipt of Stolen Property**

**Cal. Penal Code § 496(a), (c)**

**(Against All Defendants on Behalf of the Unregistered Class)**

167. Plaintiffs incorporate by reference the preceding factual allegations.

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

169. Defendants knowingly received, concealed, and withheld property—specifically, digital files embodying Plaintiffs’ copyrighted works—that had been obtained in a manner constituting theft, including unauthorized copying and distribution from shadow libraries, in violation of Cal. Penal Code § 496(a). Defendants knew or had reason to know that the property was obtained without the consent of the rightful owners and in violation of law

170. Under Cal. Penal Code § 7, “the word ‘person’ includes a corporation as well as a natural person.” Thus, Defendants are persons under Cal. Penal Code § 496(a).

171. As discussed above, Defendants unlawfully acquired copyrighted material from the internet, including by torrenting works from shadow libraries and violating websites’ terms and conditions, and used the material to develop GenAI models and products in order to generate massive profits. At no point did Defendants have consent to take/scrape this information and use it in connection with their GenAI models and products. Defendants meet the grounds for liability under Cal. Penal Code § 496(a) because each of them:

- a. Knew that the taken copyrighted material was stolen or obtained without permission, and with such knowledge;
- b. Concealed, withheld, or aided in concealing or withholding said data from their rightful owners by unlawfully manipulating (for example, removing CMI) and using the data to train their GenAI models.

1 172. Pursuant to Cal. Penal Code § 496(c), Plaintiffs, on behalf of themselves and the Class,  
2 seek actual damages, treble damages, costs of suit, and reasonable attorneys' fees.

3 **COUNT 12**

4 **Sherman Act – Conspiracy to Restrain Trade**

5 **15 U.S.C. §§ 1 & 3**

6 **(Against All Defendants on Behalf of the Non-Book Infringement Class and the Unregistered**  
7 **Class)**

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

9 174. Defendants OpenAI and Microsoft entered into a partnership, including substantial  
10 financial investments and operational support, to develop and commercialize large language models.

11 175. All textual work has value for use as training data for LLMs. This is demonstrated by  
12 the numerous licensing deals announced between generative AI companies such as OpenAI and  
13 Microsoft with licensors of textual works. This includes licenses between AI companies and licensors  
14 of registered works such as the highly publicized licensing deal between Microsoft and HarperCollins,  
15 and licenses between AI companies and licensors of unregistered textual works such as between  
16 OpenAI and Reddit.

17 176. The relevant product market is the market for registered copyright works and  
18 unregistered textual works for LLM training data, broadly defined as data that can be used in  
19 consideration for training, evaluation, validation, and actual training. The relevant geographic market is  
20 nationwide.

21 177. Defendants OpenAI and Microsoft are horizontal competitors for training data, and both  
22 compete in the markets for registered copyrighted works and unregistered text for training data.

23 178. The relevant product market is the market for copyrighted and unregistered textual  
24 works used as training data for large language models in the United States. Defendants OpenAI and  
25 Microsoft, as horizontal competitors in this market, entered into an agreement to share and exchange  
26 training data, including unlawfully acquired copyrighted works, with the purpose and effect of  
27 suppressing the price and availability of such data, foreclosing competition, and restraining trade. As a  
28 direct and proximate result, Plaintiffs and the Class suffered antitrust injury, including reduced

1 compensation for their works, diminished market opportunities, and suppression of innovation and  
2 output in the market for AI training data.

3 179. As a key part of that partnership, OpenAI and Microsoft reached an agreement and  
4 common understanding. While the arrangements were in part formalized in written documentation, the  
5 basic purpose and effect of the agreement was that OpenAI and Microsoft would cooperate to prevent  
6 the development of a fee and open market for training data.

7 180. In particular, OpenAI agreed to share training data for LLMs with Microsoft. On its part,  
8 instead of entering the market and buying copyrighted and unregistered works for its own training data,  
9 Microsoft received those same works from OpenAI, who had previously stolen them, and then  
10 provided them to Microsoft in exchange for data Microsoft had from its Bing search engine team's  
11 work on data acquisition. OpenAI, in turn, would receive training data derived from textual works  
12 scraped from the public internet for pennies on the dollar.

13 181. As a result of that agreement and understanding, OpenAI and Microsoft artificially and  
14 unreasonably restrained the market for training data for LLMs. Instead of Microsoft entering the market  
15 and purchasing copyrighted works for its own training data, it received those works from OpenAI in  
16 exchange for its own Bing data (on information and belief, comprised of both registered copyright  
17 works and unregistered works). This arrangement represents anticompetitive conduct between  
18 horizontal competitors in the market for training data for LLMs and as such limited competition in the  
19 AI training space, since Microsoft did not compete with OpenAI to acquire high-quality training data  
20 for LLMs. The agreement to "tone down the effort on training on the large models" and to be "super-  
21 careful not competing with [OpenAI]" further solidified the restraint of trade. As a direct and  
22 foreseeable result, OpenAI and Microsoft paid less for training data than they would have but for their  
23 agreement and common course of conduct.

24 182. As a direct and proximate result of Defendants' conspiracy to restrain trade, Plaintiffs  
25 and the Class have suffered antitrust injury, including but not limited to the loss of control over their  
26 registered copyrighted works and unregistered works, the unauthorized use of their intellectual  
27 property, and diminished market value of their works. The conspiracy affected Plaintiffs' ability to  
28 compete in the market, resulting in financial losses and other specific damages. By artificially

1 restraining the price of training data and limiting competition, Defendants’ actions have caused  
2 significant harm to Plaintiffs and the market as a whole.

3 183. Because Defendants are horizontal competitors in the relevant markets for registered  
4 copyrighted works as training data, and unregistered textual works for training data, Defendants’  
5 agreement to restrain trade constitutes a *per se* violation of the Sherman Act.

6 184. While the conspiracy constitutes a *per se* violation of the Sherman Act, Defendants also  
7 exploited their collective market power in the relevant market, which is the market for training data for  
8 LLMs in the United States.

9 185. Through their conspiracy, Defendants exercised and maintained market power, and did  
10 in fact suppress the market value for copyrighted works as training data for LLMs.

11 186. The purpose and effect of this restraint of trade was to restrain competition. Prices  
12 decreased, output decreased and innovation was constrained.

13 187. The conduct was not part of a legitimate joint venture and was not ancillary to another  
14 legitimate agreement.

15 188. The conspiracy and the conduct of Defendants and their agents and co-conspirators in  
16 furtherance thereof did not have procompetitive effects and were not intended to have procompetitive  
17 effects.

18 189. In the alternative, any procompetitive effects that may have resulted from the conspiracy  
19 are substantially outweighed by the anticompetitive harm alleged herein, including, but not limiting to  
20 eliminating Class members’ ability to control their works and suppressing the price of copyrighted  
21 works as training data for LLMs.

22 190. Defendants are also liable under a “quick look” analysis where one with even a  
23 rudimentary understanding of economics could conclude that the arrangements and agreements alleged  
24 would have an anticompetitive effect on Class members and the relevant market.

## 25 X. DEMAND FOR JUDGMENT

26 Wherefore, Plaintiffs request that the Court enter judgment on their behalf and on behalf of the  
27 Class defined herein, by ordering and decreeing:

- 28 a. This Action may proceed as a class action, with Plaintiffs serving as Class

1 Representatives, and with Plaintiffs' counsel as Class Counsel;

- 2 b. A declaration that Defendants have infringed Plaintiff Catherine Denial's and  
3 members of the Non-Book Infringement Class's exclusive copyrights, including  
4 but not limited to those in the Selected Infringed Works, under the Copyright Act;
- 5 c. A declaration that such infringement is willful;
- 6 d. Judgment in favor of Plaintiffs and the Class and against Defendants;
- 7 e. An award of statutory and other damages under 17 U.S.C. § 504 for Defendants'  
8 willful infringement of Plaintiff Catherine Denial's and members of the Non-Book  
9 Infringement Class's exclusive copyrights;
- 10 f. Defendants have engaged in a trust, contract, combination, or conspiracy in  
11 violation of Sections 1 and 3 of the Sherman Act, and that Plaintiffs and Class  
12 members have been damaged and injured in their business and property as a result  
13 of this violation;
- 14 g. The alleged combinations and conspiracy are *per se* violations of the Sherman Act;
- 15 h. Reasonable attorneys' fees and costs as available under 17 U.S.C. § 505, Cal. Penal  
16 Code § 502 or other applicable statute;
- 17 i. Pre- and post-judgment interest on the damages awarded to Plaintiffs and the  
18 Class, and that such interest be awarded at the highest legal rate from and after the  
19 date this class action complaint is first served on Defendants;
- 20 j. Defendants are to be jointly and severally responsible financially for the  
21 costs and expenses of a Court-approved notice program through post and  
22 media designed to give immediate notification to the Class;
- 23 k. Nominal, treble, and punitive damages, as warranted;
- 24 l. Permanent injunctive relief, including but not limited to the return,  
25 destruction, and cessation of the use of any data illegally or unlawfully  
26 acquired, or of the products dependent upon the use thereof;
- 27 m. Restitution and non-restitutionary disgorgement of all profits obtained as a  
28 result of Defendants' unjust enrichment, as well as other equitable relief as

1 provided by law;

2 n. Further relief for Plaintiffs and the Class as may be just and proper.

3 **JURY TRIAL DEMANDED**

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

6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

By: Joseph R. Saveri  
**JOSEPH SAVERI LAW FIRM, LLP**  
Joseph R. Saveri (SBN 130064)  
Cadio Zirpoli (SBN 179108)  
Christopher K.L. Young (SBN 318371)  
Holden Benon (SBN 325847)  
Aaron Cera (SBN 351163)  
Alexander Y. Zeng (SBN 360220)  
jsaveri@saverilawfirm.com  
czirpoli@saverilawfirm.com  
hbenon@saverilawfirm.com  
cyoung@saverilawfirm.com  
acera@saverilawfirm.com  
azeng@saverilawfirm.com 601 California Street,  
Suite 1505  
San Francisco, CA 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940

By: Bryan L. Clobes  
**CAFFERTY CLOBES MERIWETHER  
& SPRENGEL LLP**  
Bryan L. Clobes (*pro hac vice*)  
Mohammed A. Rathur (*pro hac vice*)  
bclobes@caffertyclobes.com  
mrathur@caffertyclobes.com  
135 South LaSalle Street, Suite 3210  
Chicago, IL 60603  
Tel: (312) 782-4880

By: Joshua Michelangelo Stein  
**BOIES SCHILLER FLEXNER LLP**  
Maxwell V. Pritt (SBN 253155)  
Joshua Michelangelo Stein (SBN 298856)  
Reed D. Forbush (SBN 347964)  
mpritt@bsflfp.com  
jstein@bsflfp.com 44 Montgomery Street, 41st  
Floor  
San Francisco, CA 94104  
Telephone: (415) 293-6800

David Boies (*pro hac vice*)  
dboies@bsflfp.com 333 Main Street  
Armonk, NY 10504  
Telephone: (914) 749-8200

Jesse Panuccio (*pro hac vice*)  
jpanuccio@bsflfp.com  
1401 New York Ave, NW  
Washington, DC 20005  
Telephone: (202) 237-2727

Evan Matthew Ezray (*pro hac vice*)  
eezray@bsflfp.com  
401 E. Las Olas Blvd., Suite 1200  
Ft. Lauderdale, FL 33301  
Telephone: (954) 356-0011

*Counsel for Individual and Representative  
Plaintiff and the Proposed Class*