

INTRODUCTION

1. This is a class action brought by textbook authors (the “Class”) whose copyrighted textbooks Defendants copied, reproduced, distributed, and used without authorization to build and train OpenAI’s generative large language model (“LLM” or “model”) known as GPT-N.

2. Without license, consent, credit, or compensation to the textbook authors, Defendants obtained Plaintiffs’ and the Class’s copyrighted textbooks from pirated “shadow library” sources, reproduced them, distributed them to others, removed copyright-management information from them, and used them to train their LLM.

3. Plaintiffs are textbook authors. Textbook authors occupy a distinct position in the market for written works, as textbooks are created, marketed, and sold for the specific purpose of teaching academic subjects to students in classroom and course settings at educational institutions. Textbooks are reviewed and selected by instructors through formal adoption processes and are typically updated, revised, and published in successive editions on a regular basis. These are commercialized educational markets that fundamentally differ from the markets for trade books and scholarly works.

4. Those distinctions are central to this litigation. The market for textbooks and the risk that AI models trained on textbooks will substitute for, dilute, or erode that market turns in part upon evidence of textbook adoption, revision economics, and customer purchasing decisions that are unique to textbook authors as compared to other categories of authors.

5. Each Plaintiff assigned exclusive publication and distribution rights under the Copyright Act to his or her respective publisher, subject to the publisher’s contractual obligation to pay royalties to the Plaintiff based on the publisher’s exploitation of those exclusive rights. Accordingly, each Plaintiff is the beneficial owner of copyright in his or her work and is entitled to institute an action for infringement of the copyright, pursuant to 17 U.S.C. § 501(b).

6. Plaintiffs assert claims for direct copyright infringement, contributory copyright infringement, vicarious copyright infringement, and distribution of works with copyright management information removed under the Digital Millennium Copyright Act, on behalf of themselves and a class of textbook authors.

JURISDICTION AND VENUE

7. This Court has subject-matter jurisdiction over this action under 28 U.S.C. §§ 1331 and 1338 because this action arises under the Copyright Act, 17 U.S.C. §§ 101 *et seq.*, and the Digital Millennium Copyright Act, 17 U.S.C. §§ 501, 1202, and 1203.

8. This Court has personal jurisdiction over Defendants. Defendants have purposefully availed themselves of this forum and the conduct giving rise to Plaintiffs' claims occurred substantially in this District.

9. OpenAI's and Microsoft's copyright infringement, contributory copyright infringement, vicarious copyright infringement, and distribution of works with copyright management information removed occurred, in substantial part, in this District. OpenAI marked, sold, and distributed, and continues to market, sell, and distribute, their GPT products within New York and to New York residents and New York-based companies.

10. Microsoft distributed and sold GPT-based products, like Bing Chat and Azure, in New York. Upon information and belief, Microsoft also assisted OpenAI's copyright infringement from New York, including from Azure datacenters located in New York, which were used to facilitate OpenAI's use and exploitation of the training dataset used for development of OpenAI's GPT models. Microsoft maintains offices and employs personnel in New York who contributed to OpenAI's infringement.

11. Venue is proper in this District under 28 U.S.C. § 1391(b)(2) because a substantial part of the events giving rise to Plaintiffs' claims occurred here.

12. Venue is also proper in this District under 28 U.S.C. § 1400(a) because Defendants or their agents reside or may be found here.

PARTIES

I. Plaintiffs

13. Plaintiff Michael Sullivan is an individual residing in Florida and is Professor Emeritus of Mathematics at Chicago State University, where he taught for more than 35 years.

14. Dr. Sullivan is the author of more than 120 mathematics textbooks and editions published by Pearson Education, Inc., W.H. Freeman & Co. and BFW Publishing Group (subsidiaries of Macmillan Publishers) and John Wiley & Sons, Inc., including the widely adopted titles *College Algebra*, *Algebra and Trigonometry*, *Precalculus*, *Trigonometry: A Unit Circle Approach*, 12th Edition (Pearson Education), 3-book *Precalculus series*, *Concepts Through Function*, 5th Edition (Pearson Education), 3-book *Precalculus Enhanced with Graphing Utilities series*, 9th Edition (Pearson Education), and *Calculus for the AP Course*, 4th Edition (BFW Publishers).

15. Dr. Sullivan has beneficial interests, copyrights, exclusive rights, royalty interests, contractual rights, or other legally protected interests in these and other works that give him standing within the class.

16. Dr. Sullivan is a past president of the Textbook & Academic Authors Association (“TAA”), a member organization whose mission is to advance the interests of textbook and academic authors. TAA’s “Michael Sullivan Lecture on Textbook & Academic Authoring” is named in his honor. He currently serves as TAA’s Treasurer.

17. Plaintiff Kenneth S. Saladin is an individual residing in Georgia and is Distinguished Professor of Biology, Emeritus, at Georgia College & State University.

18. Dr. Saladin is the author of 21 titles and editions of English-language textbooks of human anatomy and physiology published by McGraw Hill, LLC. Dr. Saladin's textbooks include *Anatomy & Physiology: The Unity of Form and Function*, *Human Anatomy*, and *Essentials of Anatomy & Physiology*.

19. Dr. Saladin has beneficial interests, copyrights, exclusive rights, royalty interests, contractual rights, or other legally protected interests in these and other works that give him standing within the class.

20. Plaintiff Zvi Bodie is an individual residing in Massachusetts and is Professor Emeritus of Finance at the Questrom School of Business at Boston University, where he taught for more than four decades.

21. Dr. Bodie is a co-author of the leading finance textbooks published in multiple editions, including *Investments and Essentials of Investments*, published by McGraw Hill, LLC, and *Financial Economics*, published by Cambridge University Press.

22. Dr. Bodie has beneficial interests, copyrights, exclusive rights, royalty interests, contractual rights, or other legally protected interests in these and other works that give him standing within the class.

23. Plaintiff Alan J. Marcus is an individual residing in Massachusetts and is the Mario J. Gabelli Endowed Professor of Finance at the Carroll School of Management at Boston College.

24. Dr. Marcus is a co-author of leading finance textbooks in multiple editions, including *Investments and Essentials of Investments* and *Fundamentals of Corporate Finance*, all published by McGraw Hill, LLC.

25. Dr. Marcus has beneficial interests, copyrights, exclusive rights, royalty interests, contractual rights, or other legally protected interests in these and other works that, give him standing within the class.

26. Kevin T. Patton is an individual residing in Missouri who worked as an anatomy & physiology (A&P) professor for several decades, teaching at the high school, community college, undergraduate, and graduate levels. Dr. Patton is President Emeritus of the Human Anatomy & Physiology Society and a founder of the HAPS Institute, a continuing education program for A&P professors. He is also a past President of the Textbook & Academic Authors Association.

27. Dr. Patton is the author of multiple successive editions of widely adopted anatomy and physiology textbooks published by Elsevier, Inc., including *Anatomy & Physiology, Structure & Function of the Body*, and *The Human Body in Health & Disease*. Dr. Patton authored his textbooks as an employee of Lion Den Inc., an S-Corporation of which he is the president and owner.

28. Lion Den Inc. is a Missouri S-Corporation solely owned by the Patton Trust dated May 13, 2005, of which Dr. Patton and his wife are both grantors and trustees.

29. Lion Den Inc., together with Dr. Patton, have beneficial interests, copyrights, exclusive rights, royalty interests, contractual rights, or other legally protected interests in these and other works that give him and his company standing within the class.

30. Dana Loewy is an individual residing in California and is Professor Emerita of Business Communication at Cal State Fullerton.

31. Dr. Loewy is the co-author of multiple successive editions of leading business communications textbooks published by Cengage Learning, Inc., including *Business Communication: Process & Product* and *Essentials of Business Communication*.

32. Dr. Loewy has beneficial interests, copyrights, exclusive rights, royalty interests, contractual rights, or other legally protected interests in these and other works that give her standing within the class.

33. Each Plaintiff assigned exclusive publication and distribution rights to his or her respective publisher via a publishing contract with the publisher. The grant of rights is subject to the publisher's contractual obligation to pay royalties to the Plaintiff based on the publisher's exploitation of those exclusive rights. Accordingly, each Plaintiff is the beneficial owner of copyright in his or her work and is entitled to institute an action for infringement of the copyright, pursuant to 17 U.S.C. § 501(b). If the publishing contract is terminated, the grant of copyright typically reverts to the Author, who then holds both legal and beneficial copyright.

34. A schedule of named Plaintiffs' representative works is attached hereto as **Exhibits A-1 to A-6**. One or more edition and/or version of each of the representative works has been registered as a literary work in the U.S. Copyright Office. Each of the registrations is valid and subsisting.

II. Defendants

35. Defendant OpenAI Foundation (f/k/a OpenAI Inc.) is a Delaware nonprofit organization with its principal place of business in San Francisco, California.

36. Defendant OpenAI Group PBC is a Delaware public benefit corporation with its principal place of business in San Francisco, California.

37. Defendant OpenAI GP, LLC is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

38. Defendant OpenAI OpCo, LLC is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

39. Defendant OpenAI Global, LLC is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

40. Defendant OAI Corporation is a Delaware corporation with its principal place of business in San Francisco, California.

41. Defendant OpenAI Startup Fund I LP is a limited partnership formed under the laws of Delaware with its principal place of business in San Francisco, California.

42. Defendant OpenAI Startup Fund GP I LLC is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

43. Defendant OpenAI Startup Fund Management LLC is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

44. Defendant Microsoft Corporation is a Washington corporation with a principal place of business in Redmond, Washington.

FACTUAL ALLEGATIONS

I. General Allegations

45. Plaintiffs, individually and on behalf of the Class, allege that OpenAI reproduced and distributed their copyrighted works without authorization in connection with sourcing content for, developing, and training OpenAI's LLMs, which are collectively referred to as "GPT-N," an abbreviation for "Generative Pre-trained Transformer" followed by a version number.

46. GPT-1 was released by OpenAI in June 2018.

47. GPT-2 was released by OpenAI in February 2019.

48. GPT-3 was released in 2020 and exclusively licensed to Microsoft the same year.

49. OpenAI further refined GPT-3 into GPT-3.5, which was released in 2022.

50. GPT-4, the successor to GPT-3.5, was released in March 2023.

51. GPT-5 was released in August 2025. GPT-5.6, the most recent member of the GPT-5 family, was released on July 9, 2026.

52. GPT-4 and the successor models underly Microsoft's Bing Chat product, offered through its Bing Internet search engine, and is integrated into its sales and marketing software, coding tools, productivity software, and cloud storage services.

53. As of June 2026, 92 percent of Fortune 500 companies use OpenAI's products.

54. All of the active versions of GPT-N are available to consumers and businesses for commercial and/or personal usage.

II. Training GPT-N on Copyrighted Books

55. Like other LLMs, such as Llama and Claude, GPT-N is a type of artificial intelligence trained on massive text datasets to understand, summarize, translate, and generate human-like language.

56. GPT-N, like other LLMs, functions as the engine behind modern chatbots by predicting probable words in a sequence based on patterns learned from billions of text sources.

57. The training of an LLM is critical since the more data an LLM has to train on, the better it is at predicting word patterns.

58. Thus, the more data a model is trained on, the more commercial viability it has.

59. OpenAI has trained their LLMs on datasets that include copyrighted works. Among the training datasets OpenAI used to train the model were Common Crawl, Books1, and Books2.

60. Common Crawl consists of raw web page data, metadata extracts, and text extracts scraped from billions of web pages. Common Crawl is widely used in training LLMs and has also been used to train Meta's Llama and Google's BERT. Common Crawl contains text from books copied from pirate sites.

61. Books1 and Books2 contain ebook files downloaded from Library Genesis, or “LibGen,” which offers a vast repository of pirated books.

62. The books aggregated by these websites have also been available in bulk via torrent systems, such as Bibliotik, a pirate torrent tracker.

63. The scale of OpenAI’s copying and distribution was massive. The Books2 dataset, contains 55 billion “tokens,” the basic units such as words, syllables, numbers, and punctuation marks, which suggests it contains over 100,000 books.

64. OpenAI’s copying occurred in multiple stages, including downloading works, copying works into computer memory, converting works into formats usable by AI systems, and copying works into the training materials used to build the GPT-N models.

65. OpenAI also made available, uploaded, and distributed copies of those works to other users worldwide, thereby further facilitating copyright infringement.

66. Separately, OpenAI developed and used software to remove copyright-management information (“CMI”), including copyright notices, titles, International Standard Book Numbers (“ISBNs”), and similar identifying information, from books it obtained from shadow libraries. OpenAI did this to facilitate training and to conceal their use of copyrighted works.

67. Specifically, OpenAI developed and deployed a software script that removed CMI from the beginning and end of works obtained from shadow libraries, including ISBNs, copyright notices, the © symbol, statements of “all rights reserved,” and Digital Object Identifiers (“DOIs”), as well as language indicating that unauthorized copying constituted infringement.

68. Beyond stripping CMI, OpenAI took further steps to conceal their use of copyrighted works, including by addressing instances in which the GPT-N models, in response to user prompts, generated answers acknowledging that their training data might include pirated,

illegal, or copyrighted material; OpenAI's personnel introduced curated fine-tuning samples designed to suppress such answers and to reduce the risk that users or the public would learn that OpenAI had trained GPT-N on infringing material.

69. The GPT-N models generate outputs consisting of up to 90 characters of copyrighted textbooks, along with the schematic structure of those works. For example, when prompted with questions about Dr. Sullivan's textbooks, ChatGPT returned detailed, accurate summaries of them, even reproducing specific examples from his textbooks.

70. Corporate clients have integrated the GPT-N models into their principal commercial products and services, including Microsoft's Bing Chat product offered through its Bing Internet search engine, and OpenAI has derived substantial commercial benefit from models trained on the copied works.

III. Microsoft's Relationship with OpenAI

71. OpenAI has relied on Microsoft's financial and technical support to train its LLMs.

72. Microsoft released Azure, a cloud computing system, for use in 2010.

73. In 2020 OpenAI partnered with Microsoft to use Azure as their exclusive cloud provider to train OpenAI's AI models.

74. Microsoft built and made available new infrastructure in Azure to train extremely large AI models. Microsoft and OpenAI collaborated to build the supercomputer hosted in Azure that was designed specifically to train OpenAI's AI models. Without Microsoft's computer system, OpenAI would not have been able to execute and profit from their copyright infringement.

75. According to Satya Nadella, Microsoft's Chief Executive Officer ("CEO"), "[OpenAI] do[es] the foundation models, and we do a lot of work around them, including the

tooling around responsible AI and AI safety.”¹ Microsoft has taken credit for its role in the training process of GPT-N, with Mr. Nadella stating that Microsoft “buil[t] the infrastructure to train [OpenAI’s] models.”²

76. Microsoft exchanges know-how and data with OpenAI and develops, maintains, troubleshoots, supports, and controls the computing system used by OpenAI to train their LLMs.

77. By virtue of its enmeshment in OpenAI’s training process of their LLMs, Microsoft would have known that OpenAI’s training data was scraped from the internet and included a massive number of copyrighted books. Microsoft made unlicensed copies and assisted OpenAI in making unlicensed copies of copyrighted material, including Plaintiffs’ works and other copyrighted textbooks to train the GPT models.

78. Microsoft had an incentive to ensure that OpenAI’s LLMs were trained on the quality of data necessary to ensure their success.

79. OpenAI and Microsoft are such close partners that Microsoft hired OpenAI CEO Sam Altman when Mr. Altman was terminated in November 2023. Thereafter, OpenAI reinstated Mr. Altman as CEO and granted Microsoft a nonvoting seat on the board of OpenAI, Inc.

80. Microsoft has also commercialized OpenAI’s technology, including it in Bing Chat, its “Copilot” application offerings, and in the GitHub Copilot AI coding editor. ChatGPT integrated a “Browse with Bing” feature in its paid offering.

81. Microsoft has a strategic partnership with OpenAI through 2030.

¹ “OpenAI’s GPT-4 helped Microsoft CEO Satya Nadella live a childhood dream,” available at <https://www.windowcentral.com/microsoft/gpt-4-helped-microsoft-ceo-satya-nadella-live-a-childhood-dream-leading-to-a-eureka-moment> (last accessed July 24, 2026).

² Satya Nadella July 18, 2023 Microsoft Inspire Keynote Address, available at <https://msftstories.thesourcemediaassets.com/2023/07/Microsoft-Inspire-2023.pdf> (last accessed July 24, 2026).

IV. Textbooks and Textbook Authors

82. Defendants' conduct affected works across a broad range of categories, including fiction, nonfiction, children's books, poetry, memoirs, dramatic works, scholarly works, as well as textbooks.

83. The putative class consists of all textbook authors holding legal or beneficial copyrights in works reproduced, distributed, or otherwise used by Defendants without authorization.

84. On information and belief, each copied work is an original work, comprises copyrightable subject matter, is registered with the U.S. Copyright Office, and remains subject to valid copyrights in which class members have a legal or beneficial interest.

85. Textbooks are written works created for the express purpose of teaching academic subjects to primary school, high school, college, graduate, professional, and continuing education students and are marketed, offered, or selected for assignment to students in classroom or course settings.

86. The primary goal of textbooks is to teach the content of a field or subfield in a knowledgeable, coherent, and approachable way that serves the purposes of instructors and students in that field or subfield. Over decades of writing these textbooks, each author develops a unique voice that becomes their distinctive style. A particular textbook is selected and adopted not only because of the title or subject matter but because the author has a style that works for the intended learners.

87. For example, Dr. Sullivan's *Precalculus*, now in its 12th Edition, teaches the content of the undergraduate precalculus course, including graphs, functions, and analytic geometry, in a way that is accessible to students and that prepares those students to go on to study calculus.

88. Dr. Saladin's *Anatomy and Physiology: The Unity of Form and Function*, now in its 10th Edition, has achieved its large market share mainly on the strength of a unique style and pedagogic organization not seen in competing textbooks on the same subject.

89. Drs. Bodie's and Marcus's *Essentials of Investments*, now in its 13th Edition, presents investment analysis and investment theory to undergraduate business students in a way that is consistent with the approach of the CFA (Chartered Financial Analyst) Institute.

90. Dr. Patton's *Anatomy & Physiology* is designed as an accessible course book, especially for teaching nursing students or others entering the biological sciences, and for instructors desiring a more conversational, narrative approach to the subject matter.

91. Dr. Loewy's *Essentials of Business Communication*, now in its 13th edition, teaches best practices in business communication supported by the latest research, with a practical focus on current technology and transferable workplace skills.

92. Most textbook authors are experienced classroom instructors, skilled writers, and subject matter experts in their fields. They each bring their unique combination of knowledge, experience, and skills to writing their textbooks.

93. Textbooks in any particular academic area cover the field in a variety of ways, reflecting the unique backgrounds, interests, and styles of their authors, as well as the needs of their intended audience of instructors and students.

94. The subject matter that a textbook conveys, whether trigonometric functions, the respiratory system, or the operation of security markets, is not itself protected by copyright. What is protected, and what gives a textbook its value, is the author's selection, organization, and sequencing of material, the examples and explanations that the author develops to effectively convey the information to the student, and the author's voice and pedagogical style.

95. Unlike trade books, which are seldom revised, textbooks typically have regular revision cycles to keep their treatment of the subject matter up to date with developments in the field, their pedagogy current, and to stay aligned with instructor preferences, student needs, and adoption standards. Each revision requires a substantial investment of time, research, and resources by the author. One result of illegal training on pirated textbooks, and market substitution of students who turn to artificial intelligence alternatives, is that students can receive inferior, incomplete, or out-of-date guidance that *appears to be* authoritative. Searches on ChatGPT for the named Plaintiffs and their publications resulted in exactly such responses, providing feedback to users that purports to be “in the style of” or “taking the approach of” specifically named textbook authors. For example, when asked to explain the first chapter of Dr. Sullivan’s *Precalculus: Enhanced with Graphing Utilities*, 6th Edition, ChatGPT stated: “I’ll use the same sort of progression Sullivan uses”

96. Many textbooks, including those authored by Plaintiffs, have lives that extend over many editions and many decades. These textbooks become a central part of the authors’ life work, commitments, reputations, and livelihoods.

97. Although different textbooks in any field have different emphases, viewpoints, and styles, successful textbooks share a reputation in their fields, and in particular among the adopters and instructors who select them for use in the classrooms, for accessibility, reliability, and effectiveness.

98. Textbooks play an important role in shaping the academic fields and subfields that they cover, and in shaping the education of the students who use them.

99. Because of the importance of textbooks in academia, many textbook authors establish their professional reputation not only on the basis of their teaching and research, but also on the quality, impact, and adoptions of their textbooks.

V. The Textbook Market

100. The markets for textbooks, on the one hand, and trade books (both fiction and non-fiction), on the other, are fundamentally different. The trade book market consists primarily of people who choose to purchase books for personal reasons, *e.g.*, for their own edification, curiosity, or entertainment. A casual reader looking for a novel by George R.R. Martin or a popular history book by David McCullough will not accept a substitute by another author or generated by an LLM.

101. In the textbook market, by contrast, the final purchaser is *not* the person who selects the book. The market for textbooks is driven by adoptions, *i.e.*, instructors, schools, or educational systems who consider the textbook options available in a specific subject area and then “adopt” the one they think best suits the needs of their instructors and students. The textbooks are then assigned to students in specific courses.

102. Purchasing decisions are fundamentally different in the textbook market than in the trade book market because the person who *chooses* the textbook is not the same person who is *required* to purchase it. The purchaser at the elementary or secondary school level is typically a school system or individual school required to make the purchase and is continually looking for ways to cut costs from its budget. In the higher education space, the purchaser is typically the student assigned the textbook by their instructor. In either situation, the purchaser has a powerful incentive to avoid the cost of purchasing a textbook if a “free” version of that same book, or a “good enough” substitute can be found via an LLM.

VI. Defendants' Use of Textbooks to Train Their LLMs Threatens to Displace the Textbook Market

103. Because textbooks cover academic fields and subfields in engaging, effective and up-to-date ways, they are uniquely valuable in the training of LLMs.

104. On information and belief, textbooks constitute a large proportion of the books that Defendants used to train their GPT-N models.

105. Defendants' GPT-N models trained on sets of textbooks in various fields are capable of feeding back to users comprehensive information about those fields, often with detailed references to specific authors and their organization of and approach to the subject matter.

106. Textbooks are required purchases by students (in higher education) and schools (in K-12 settings). Students, instructors, and schools feel strong pressures to reduce the cost of educational materials. As a result, textbooks are uniquely susceptible to displacement by LLMs trained on textbooks in that field.

107. The GPT-N models can generate outputs that substitute for copyrighted works, including outlines, summaries, low-quality alternative versions, imitations, and derivative works. This impacts the market for textbooks in particular, insofar as textbooks are designed to teach and impart knowledge and structures for learning.

108. As LLMs become more powerful and widespread, and if they are allowed to continually update their content by ingesting new editions of textbooks as they come out, instructors, schools, and adoption and curriculum committees may consider dropping textbook adoptions altogether in favor of substitutes that rely on the use of LLM models. Alternately, the availability of LLM substitutes may have the effect of eroding the pricing for textbooks.

109. LLMs threaten the textbook market by destroying the incentives for textbook authors to continue writing and revising high quality textbooks, to the detriment of the textbook

authors on whose work the LLMs rely and the student learners whose education may be compromised by their reliance on outputs.

110. GPT-N and other generative AI agents threaten to displace human-authored textbooks with cheaper or no-cost AI-generated textbooks or textbook equivalents, at the expense of losing the valuable authorial voice, research, expertise, and perspective provided by textbook authors.

111. On information and belief, licensing markets exist, or could be developed, that would permit Defendants to make use of Plaintiffs' textbooks in a controlled context and in exchange for fair compensation. Defendants bypassed such licensing efforts by obtaining Plaintiffs' textbooks without license, thus threatening to displace the very sources on which they relied.

112. Using copyrighted textbooks to train an LLM harms the market for those works by helping to enable the rapid generation of countless works that compete with the originals, even if those works themselves are not necessarily infringing. LLMs can generate massive amounts of text in significantly less time that it would take to write that text, and using a fraction of the creativity.

VII. Defendants Willfully Engaged in Copyright Infringement

113. A market exists for licensing copyrighted works for AI training. AI companies could have tried to negotiate agreements with publishers, media companies, academic institutions, or individual authors for access to books, articles, and journals for model development and training. OpenAI did not take these actions with respect to the textbooks it copied and distributed.

114. On information and belief, when Defendants used massive numbers of books without permission or compensation to their authors, they were aware that such use would likely displace, dilute, and erode the market for the works used in the training.

115. On information and belief, Defendants’ unauthorized use displaced legitimate sales and licenses, deprived authors and publishers of revenues, usurped the market for licensing copyrighted works for AI training, and has eroded the price that a textbook can command.

116. Throughout, Defendants acted willfully and with knowledge that their conduct was unlawful.

117. OpenAI cross-referenced pirated titles in their LibGen collection against books available for license in order to decide whether to pursue licenses.

118. OpenAI has admitted that they has trained their LLMs on “large, publicly available datasets that include copyrighted works.”³

119. OpenAI has admitted that training LLMs “require[s] large amounts of data,” and that “analyzing large corpora” of data “necessarily involves first making copies of the data to be analyzed.”⁴

120. OpenAI has admitted that, if it was prevented from using copyrighted works in their LLMs’ training, it would “lead to significant reductions in model quality.”⁵

121. Microsoft willfully made unlicensed copies and assisted OpenAI in making unlicensed copies of copyrighted material, including Plaintiffs’ works and other copyrighted textbooks to train the GPT models.

³ OpenAI, Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation, U.S. Patent and Trademark Office Dkt. No. PTO-C-2019-0038, at 1 (2019), available at https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf (last accessed July 22, 2026).

⁴ *Id.* at 2.

⁵ *Id.* at 7 n.32.

VIII. Plaintiffs' Claims Are Timely

122. Defendants do not disclose or publicize with specificity what datasets GPT-1, GPT-2, GPT-3, GPT-3.5, GPT-4, GPT-4.5, GPT-5, GPT-5.4, GPT-5.5, GPT-5.6 or their successors were trained on. Therefore, Plaintiffs and the members of the Class did not know, and in the exercise of reasonable diligence could not have known, that OpenAI had copied their textbooks, specifically, until the publishing of Alex Reisner's articles "Revealed: The Authors Whose Pirated Books Are Powering Generative AI" in The Atlantic on August 19, 2023, followed by "These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech" on September 25, 2023, first publicly disclosed and made identifiable Defendants' copying of copyrighted textbooks to train GPT-N.

123. This action is therefore brought within three years after Plaintiffs' claims accrued under 17 U.S.C. § 507(b), and Defendants' infringing conduct has in any event continued.

CLASS ACTION ALLEGATIONS

124. Plaintiffs bring this action under Federal Rules of Civil Procedure 23(a), 23(b)(2), and 23(b)(3) on behalf of the following Class:

All textbook authors who hold legal or beneficial ownership of a United States copyright in a textbook that OpenAI or Microsoft downloaded, copied, reproduced, uploaded, otherwise distributed, and/or otherwise used to train their LLM models, and that was registered with the United States Copyright Office (i) within five years of the work's first publication and before OpenAI's copying, or (ii) within three months of the work's first publication.

125. For the purposes of the Class, a textbook is defined as a book in any medium, including printed and digital, created for the express purpose of teaching academic subject matter to primary school, secondary school, college, university, graduate, professional, or continuing education students and marketed, offered, or selected for educational instruction in a classroom or course setting.

126. Excluded from the Class are Defendants; their officers, directors, employees, parents, subsidiaries, and affiliates; all governmental entities; and the judge and chambers staff assigned to this case and their immediate families.

127. The Class is so numerous that joinder of all members is impracticable. On information and belief, textbook authors number in at least the thousands, and textbooks constitute approximately twenty-five percent of the works within the broader population of works Defendants copied.

128. Questions of law and fact common to the Class predominate over any individual questions, including whether Defendants reproduced the Class's textbooks, whether Defendants distributed them, whether Defendants removed copyright-management information, whether Defendants' conduct was willful, whether such use infringed the Class's exclusive rights, whether Defendants' use harmed the market for textbooks through substitution or dilution, or price erosion, and the appropriate relief.

129. Plaintiffs' claims are typical of the Class's claims because Plaintiffs and the Class are textbook authors whose registered textbook works were copied, distributed, and used by Defendants in the same manner and who suffered the same type of harm.

130. Plaintiffs will fairly and adequately protect the interests of the Class. They are textbook authors with no conflicts with other Class members, and they have retained counsel experienced in copyright and class-action litigation, including attorneys at Slarskey LLC and Archstone Law P.C. who have represented textbook authors for many years.

131. A class action is superior to other available methods for fairly and efficiently adjudicating this controversy, and Defendants have acted on grounds generally applicable to the Class, making final relief appropriate for the Class as a whole.

132. Plaintiffs' counsel, Slarskey LLC, Archstone Law P.C., and Gaw Poe LLP are qualified and experienced in representing textbook authors and litigating class actions and should be appointed Class Counsel under Federal Rule of Civil Procedure 23(g).

CAUSES OF ACTION

COUNT I: Direct Copyright Infringement (17 U.S.C. §§ 106(1), 106(2), 106(3), 501) against OpenAI

133. Plaintiffs incorporate by reference each of the foregoing paragraphs.

134. Plaintiffs and the Class own or hold beneficial interests in the exclusive rights in their registered textbook works under 17 U.S.C. § 106, including the rights of reproduction and distribution, and preparation of derivative works.

135. OpenAI reproduced Plaintiffs' and the Class's Works without authorization by downloading them from shadow libraries, copying them into datasets and computer memory, converting them into training-ready formats, and copying them into the materials used to train their LLMs and ChatGPT.

136. OpenAI further distributed Plaintiffs' and the Class's Works without authorization by knowingly distributing generative outputs that reproduce or are substantially similar to protected expression from Plaintiffs' and the Class's Works.

137. OpenAI's infringement was willful. Plaintiffs and the Class are entitled to statutory damages, or at their election actual damages and OpenAI's profits, together with injunctive relief, costs, and attorneys' fees under 17 U.S.C. §§ 504, 505.

COUNT II: Vicarious Copyright Infringement against Microsoft

138. Plaintiffs incorporate by reference each of the foregoing paragraphs.

139. At all relevant times, Microsoft possessed the right and ability to supervise, control, and stop or limit OpenAI's infringing conduct.

140. Microsoft designed, built, owned, and operated the Azure supercomputing infrastructure on which OpenAI copied, stored, and processed Plaintiffs' and the Class's Works and trained the accused models.

141. Microsoft controlled the deployment of the accused models within its own products and services, including Copilot, Bing, and Bing Chat, and thereby had the right and ability to determine whether, and on what terms, the infringing models and their infringing outputs would be made available to the public.

142. Despite its right and ability to supervise and control the infringing activity, Microsoft declined to exercise that authority to stop or meaningfully limit OpenAI's infringement of Plaintiffs' and the Class's Works.

143. Microsoft derived a direct financial benefit from OpenAI's infringement of Plaintiffs' and the Class's Works as it used such works to enhance its own products and services, including Copilot, Bing, and Bing Chat.

144. Microsoft's conduct was willful within the meaning of 17 U.S.C. § 504(c)(2) in that Microsoft knew or recklessly disregarded that OpenAI's conduct infringed Plaintiffs' and the Class's copyrights.

145. As a direct and proximate result of the infringement for which Microsoft is vicariously liable, Plaintiffs and the Class have been injured and are entitled to the relief set forth below, including statutory damages, or at their election actual damages and Defendants' profits, together with injunctive relief, costs, and attorneys' fees.

**COUNT III: Contributory Copyright Infringement
against all Defendants**

146. Plaintiffs incorporate by reference each of the foregoing paragraphs.

147. By uploading and making available Plaintiffs' and the Class's works to other users, Defendants induced, caused, and materially contributed to the reproduction and distribution of those works by thousands of third parties.

148. Defendants knew or had reason to know that their conduct furthered the infringement of Plaintiffs' and the Class's Works, and Defendants could have avoided distributing the works but chose not to. Defendants are liable for contributory copyright infringement, and their conduct was willful.

149. Plaintiffs and the Class are entitled to statutory damages, or at their election actual damages and Defendants' profits, together with injunctive relief, costs, and attorneys' fees.

**COUNT IV: Distribution of Works with Copyright Management Information Removed
(17 U.S.C. § 1202(b)) against all Defendants**

150. Plaintiffs incorporate by reference each of the foregoing paragraphs.

151. Plaintiffs' and the Class's textbook works contained copyright-management information, including copyright notices, the © symbol, titles, ISBNs, and the names of the authors and rights holders.

152. Defendants developed and deployed software to remove that copyright-management information from the works they obtained from shadow libraries, and did so intentionally, knowing or having reasonable grounds to know that doing so would conceal and facilitate infringement, in violation of 17 U.S.C. § 1202(b).

153. Defendants distributed copies of the works to each other and through chatbot output to consumers, but also through the delivery of enterprise chatbots and other LLM products and services to corporate customers and partners, knowing CMI had been removed from the Plaintiffs' and Class's Works without authority of the copyright owner or the law.

154. OpenAI transfers ownership of the infringing output to their users by assigning all of their rights in any output to their users.

155. Defendants knew or had reason to know the distribution would induce, enable, facilitate, or conceal infringement.

156. Plaintiffs and the Class have been harmed by Defendants' unauthorized distribution of Plaintiffs' and the Class's Works with CMI removed, which has obscured Plaintiffs' and the Class's ownership rights and facilitated further unauthorized use of their works.

157. As a result, Plaintiffs and the Class are entitled to statutory or actual damages under 17 U.S.C. § 1203, together with injunctive relief, costs, and attorneys' fees.

PRAYER FOR RELIEF

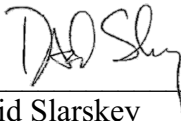
WHEREFORE, Plaintiffs, individually and on behalf of the Class, respectfully request that the Court enter judgment and grant relief as follows:

- a. Certifying this action as a class action, appointing Plaintiffs as Class Representatives, and appointing Slarskey LLC, Archstone Law P.C., and Gaw Poe LLP as Class Counsel;
- b. Entering judgment that Defendants infringed Plaintiffs' and the Class's copyrights and violated 17 U.S.C. § 1202(b);
- c. Awarding statutory damages, or at Plaintiffs' and the Class's election, actual damages and Defendants' profits, under 17 U.S.C. §§ 504 and 1203;
- d. Awarding declaratory and permanent injunctive relief, including relief restraining further infringement and requiring restoration of copyright-management information under 17 U.S.C. § 1203(b)(1);
- e. Awarding pre- and post-judgment interest, costs, and reasonable attorneys' fees as permitted by law; and

- f. Granting such other and further relief as the Court deems just and proper. Plaintiffs demand trial by jury on all claims and issues so triable.

Dated: August 14, 2026
New York, New York

SLARSKEY LLC

By: 
David Slarskey
Adam Hollander
Deepa Devanathan
Phillip Gasperetti
825 Third Avenue, 32nd Floor
New York, New York 10022
dslarskey@slarskey.com
ahollander@slarskey.com
ddevanathan@slarskey.com
pgasperetti@slarskey.com
Tel: (212) 658-0661

ARCHSTONE LAW GROUP, P.C.
Brenda Ulrich (*pro hac vice pending*)
Riverside Center
275 Grove Street, Suite 2-400
Newton, MA 02446
bulrich@archstonelaw.com
zrubin@archstonelaw.com
Tel: (781) 314-0100

GAW POE LLP
Christopher Wimmer (*pro hac vice pending*)
Randolph Gaw (*pro hac vice pending*)
One Embarcadero Suite 1200
San Francisco, CA 94111
cwimmer@gawpoe.com
rgaw@gawpoe.com
Tel: (415) 766-7451

Counsel for Plaintiffs and the Proposed Class