UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF MISSISSIPPI
NORTHERN DIVISION

BOONE NEWSMEDIA, INC. PLAINTIFF

v.

MICROSOFT CORPORATION; OPENAI, INC.; OPENAI LP; OPENAI GP, LLC; OPENAI, LLC; OPENAI OPCO, LLC; OPENAI GLOBAL, LLC; OAI CORPORATION, LLC; OPENAI HOLDINGS, LLC; OPENAI FOUNDATION; and OPENAI GROUP PBC

Civil No.: 3:26-cv-00950-DPJ-LGI

DEFENDANTS

COMPLAINT FOR DAMAGES

Plaintiff Boone Newsmedia, Inc. ("Boone" or "Plaintiff"), for its complaint against Defendants Microsoft Corporation ("Microsoft") and OpenAI, Inc.; OpenAI LP; OpenAI GP, LLC; OpenAI, LLC; OpenAI OpCo, LLC; OpenAI Global, LLC; OAI Corporation, LLC; OpenAI Holdings, LLC; OpenAI Foundation and OpenAI Group PBC (collectively "OpenAI" and, with Microsoft, "Defendants"), alleges as follows on information and belief:

INTRODUCTION

1. This is an action for preliminary and permanent injunctive relief and federal copyright infringement under 17 U.S.C. §§ 101 *et seq*. and violation of the Digital Millennium Copyright Act under 17 U.S.C. § 1202(b). Plaintiff brings this action to prevent the continuing and irreparable harm that it is suffering, and will continue to suffer, due to Defendants' willful infringement of Plaintiff's intellectual property.

2. This lawsuit arises from Defendants' systematic and willful theft of tens of thousands of copyrighted articles belonging to Plaintiff, which owns and operates local and regional newspaper outlets around the country and has spent decades investing in the journalists, editors, and infrastructure required to produce the trusted, original reporting on which its communities depend. Without permission and without any compensation to Plaintiff, Defendants scraped, copied, and ingested that content to build and commercialize their generative artificial intelligence ("GenAI") products, including ChatGPT and Microsoft Copilot. Those products have generated hundreds of billions of dollars (and counting) in market value for Defendants. Not a cent has gone to Plaintiff whose work made it possible.

3. Using automated systems, Defendants systematically and secretly crawled Plaintiff's websites, including content behind paywalls and other access restrictions, and copied Plaintiff's articles, stories, and other original works onto their own servers without authorization. As part of that process, Defendants' systems stripped from Plaintiff's works all copyright

management information ("CMI") embedded in and associated with those works, such as author credits, publication names, copyright notices, and terms of use information, that establish ownership and signal that a work is protected. That CMI-stripping, an instrumental part of Defendants' ingestion pipeline, helped sever the link between the copied content and its rightful owners and authorizations. The scraped, stripped content was then used to train Defendants' large language models ("LLMs"), which have "memorized" that material and likely reproduced it, verbatim or near-verbatim, in response to user prompts for years. And because Defendants' models must be continuously updated with new material to remain current and commercially viable, these processes have been repeated over and over and over again.

4. Defendants' conduct was willful and deliberate. OpenAI's founder, Sam Altman, acknowledged as much in testimony before the British House of Lords, conceding that it would be "*impossible to train today's leading AI models without using copyrighted materials*." Defendants made deliberate engineering choices to copy Plaintiff's content and strip its CMI, knowing it would obscure the origins of the works they were taking and impair Plaintiff's ability to detect and prove the theft. And even as litigation from other publishers mounted, and as courts began to recognize the validity of these claims, Defendants pressed forward undeterred. Defendants' data scraping processes, their products' storage and reproduction of "memorized" content, and therefore their violations of the Copyright Act and the Digital Millennium Copyright Act, continue to this day.

5. On the backs of Plaintiff, Defendants built some of the most valuable businesses in human history. OpenAI, once styled as a nonprofit, now commands a valuation worth nearly one trillion dollars. Microsoft's deployment of its Copilot product has added hundreds of billions of dollars to its market capitalization. These are not the fruits of Defendants' ingenuity alone. Plaintiff's journalism was essential to the Defendants' explosive growth, and unless Defendants

are held accountable for stealing, stripping, and misusing Plaintiff's content, the AI boom Defendants orchestrated and benefit from will be a death knell for local journalism—which remains the most trusted news sources in America.

6. Plaintiff has spent a great deal of dollars to sustain this work. Defendants helped themselves to all of it—without providing a cent of compensation.

7. The U.S. Constitution has, since the nation's founding, charged Congress with protecting authors' and Plaintiff's exclusive rights in their work. Congress has exercised that authority to implement robust protections against copyright theft, including through enacting the Copyright Act and the Digital Millennium Copyright Act, and authorizing substantial penalties for willful violations of both. Defendants invoke those protections vigorously for their own products, shielding their code, their models, and their systems behind licenses, paywalls, and legal threats. In bringing this action, Plaintiff seeks to hold Defendants to the same standard they insist upon for themselves.

JURISDICTION AND VENUE

8. This Court has subject matter jurisdiction over this action pursuant to 17 U.S.C. § 501, and 28 U.S.C. §§ 1331 and 1338 due to the federal causes of action alleged herein.

9. Venue is proper under 28 U.S.C. § 1400(a) because Defendants or their agents reside or may be found in this District, through the infringing and unlawful activities—as well as Defendants' sales and monetization of such activity—that occurred in this District. Venue is also proper under 28 U.S.C. § 1391 because a substantial part of the events giving rise to Plaintiff's claims occurred in this District, including the marketing, sales, and licensing of Defendants' GenAI products built on the infringement of Plaintiff's intellectual property within this District.

THE PARTIES

11. Plaintiff Boone Newsmedia, Inc. is a corporate citizen of Alabama. It is an Alabama corporation headquartered in Selma, Alabama. Boone owns and/or manages dozens of newspapers along with websites, shopping guides and magazines in communities in Alabama, Georgia, Michigan, Minnesota, Mississippi, Virginia, and North Carolina.

12. Boone owns and operates the following companies, which are each listed with their corresponding publication:

Clanton Newspapers Inc. - https://www.clantonadvertiser.com/
Clanton Newspapers Inc. - https://www.peachlivingmagazine.com/
Greenville Newspapers LLC - https://www.greenvilleadvocate.com/
Greenville Newspapers LLC - https://www.lowndessignal.com/
Greenville Newspapers LLC - https://www.luvernejournal.com/
Selma Newspapers Inc. - https://www.selmatimesjournal.com/
Shelby County Newspapers Inc. - https://www.shelbycountyreporter.com/
Shelby County Newspapers Inc. - https://www.shelbyliving.com/
Shelby County Newspapers Inc. - https://www.hooversmagazine.com/
Shelby County Newspapers Inc. - https://www.mountainbrookmagazine.com/
Shelby County Newspapers Inc. - https://www.vestaviahillsmagazine.com/
Shelby County Newspapers Inc. - https://www.homewoodlife.com/
Shelby County Newspapers Inc. - https://www.leedslivingmagazine.com/
Shelby County Newspapers Inc. - https://www.moodylivingmagazine.com/
Shelby County Newspapers Inc. - https://www.trussvillelivingmag.com/
Troy Publications Inc. - https://www.troymessenger.com/
Valley Newsmedia LLC - https://www.valleytimes-news.com/
LaGrange Newsmedia LLC - https://www.lagrangenews.com/
Owosso Newsmedia LLC - https://www.argus-press.com/
Albert Lea Newspapers Inc. - https://www.albertleatribune.com/
Austin Newspapers Inc. - https://www.austindailyherald.com/
Natchez Newspapers Inc. - https://www.natchezdemocrat.com/
Vicksburg Newsmedia LLC - https://www.vicksburgpost.com/
Outer Banks Newsmedia LLC - https://www.thecoastlandtimes.com/
Roanoke-Chowan Publications LLC - https://www.roanoke-chowannewsherald.com/
Roanoke-Chowan Publications LLC - https://www.gatescountyindex.com/
Washington Newsmedia LLC - https://www.thewashingtondailynews.com/
Smithfield Newsmedia LLC - https://www.smithfieldtimes.com/
Suffolk Publications LLC - https://www.suffolknewsherald.com/
Tidewater Publications LLC - https://www.thetidewaternews.com/
Tidewater Publications LLC - https://www.windsorweekly.com/

17. Defendant Microsoft Corporation is a corporate citizen of Washington. It is a Washington corporation with a principal place of business and headquarters in Redmond, Washington. Microsoft has invested at least $13 billion in OpenAI Global, LLC in exchange for which Microsoft will receive 75% of that company's profits until its investment is repaid, after which Microsoft will own a 49% stake in that company.

18. Microsoft has described its relationship with the OpenAI Defendants as a "partnership." This partnership has included contributing and operating the cloud computing services used to copy Plaintiff's works and train the OpenAI Defendants' GenAI models. It has also included, upon information and belief, substantial technical collaboration on the creation of those models. Microsoft possesses copies of, or obtains preferential access to, the OpenAI Defendants' latest GenAI models that have been trained on and embody unauthorized copies of Plaintiff's works. Microsoft uses these models to provide infringing content to users of its products and online services. And that has proven lucrative: Microsoft's 2025 end-of-year investor report touted over $75 billion in revenue from its Azure cloud computing platform.[1]

19. The OpenAI Defendants consist of a web of interrelated entities.

20. Defendant OpenAI, Inc. is a corporate citizen of Delaware and California. It is a Delaware nonprofit corporation with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI, Inc. was formed in December 2015. At least until October 28, 2025, OpenAI, Inc. indirectly owned and controlled all other OpenAI entities and was directly involved in perpetrating the mass infringement and other unlawful conduct alleged here.

21. Defendant OpenAI LP is a corporate citizen of Delaware and California. It is a Delaware limited partnership with its principal place of business located at 3180 18th Street, San

[1] Microsoft Corp, 2025 Annual Report, https://www.microsoft.com/investor/reports/ar25/index.html.

Francisco, California. OpenAI LP was formed in 2019. OpenAI LP is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit and is controlled by OpenAI, Inc. OpenAI LP was directly involved in perpetrating the mass infringement and commercial exploitation of Plaintiff's works alleged here.

22. Defendant OpenAI GP, LLC is a corporate citizen of Delaware and California. It is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI GP, LLC is the general partner of OpenAI LP, and it manages and operates the day-to-day business and affairs of OpenAI LP. OpenAI GP, LLC is wholly owned and controlled by OpenAI, Inc. OpenAI, Inc. uses OpenAI GP, LLC to control OpenAI LP and OpenAI Global, LLC. OpenAI GP, LLC was involved in perpetrating the mass infringement and unlawful exploitation of Plaintiff's works alleged here through its direction and control of OpenAI LP and OpenAI Global, LLC.

23. Defendant OpenAI, LLC is a corporate citizen of Delaware and California. It is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI, LLC was formed in September 2020. OpenAI, LLC owns, sells, licenses, and monetizes a number of OpenAI's offerings, including ChatGPT, ChatGPT Enterprise, and OpenAI's API tools, all of which were built on OpenAI's mass infringement and unlawful exploitation of Plaintiff's works. Upon information and belief, OpenAI, LLC is owned and controlled by both OpenAI, Inc. and Microsoft Corporation through OpenAI Global, LLC and OpenAI OpCo, LLC.

24. Defendant OpenAI OpCo, LLC is a corporate citizen of Delaware and California. It is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OpenAI OpCo, LLC is a wholly owned subsidiary of OpenAI,

Inc. and has facilitated and directed OpenAI's mass infringement and unlawful exploitation of Plaintiff's works through its management and direction of OpenAI, LLC.

25. Defendant OpenAI Global, LLC is a corporate citizen of Delaware and California. It is a Delaware limited liability company formed in December 2022. OpenAI Global, LLC has a principal place of business located at 3180 18th Street, San Francisco, California. Microsoft Corporation has a minority stake in OpenAI Global LLC, and OpenAI, Inc. has a majority stake in OpenAI Global, LLC, indirectly through OpenAI Holdings, LLC and OAI Corporation, LLC. OpenAI Global, LLC was and is involved in the unlawful conduct alleged herein through its ownership, control, and direction of OpenAI, LLC.

26. Defendant OAI Corporation, LLC is a corporate citizen of Delaware and California. It is a Delaware limited liability company with a principal place of business located at 3180 18th Street, San Francisco, California. OAI Corporation, LLC's sole member is OpenAI Holdings, LLC. OAI Corporation, LLC was and is involved in the unlawful conduct alleged herein through its ownership, control, and direction of OpenAI Global, LLC and OpenAI, LLC.

27. Defendant OpenAI Holdings, LLC is a corporate citizen of Delaware and California. It is a Delaware limited liability company, whose sole members are OpenAI, Inc. and Aestas, LLC, whose sole member, in turn, is Aestas Management Company, LLC. Aestas Management Company, LLC is a Delaware shell company formed for the purpose of executing a $495 million capital raise for OpenAI.

28. On or around October 28, 2025, OpenAI announced an "updated structure."[2] After this announcement, the non-profit entity formerly known as OpenAI, Inc. is "now the OpenAI Foundation," and the "for-profit" business established in 2019 is "now a public benefit corporation,

[2] OpenAI, *Our structure*, https://openai.com/our-structure.

called OpenAI Group PBC."[3] OpenAI's website notes that the "OpenAI Foundation continues to control the OpenAI Group."[4]

29. The OpenAI Foundation is a Delaware foundation with a principal office in San Francisco, California and OpenAI Group PBC is a Delaware public benefit corporation also with a principal place of business in San Francisco, California.

FACTUAL ALLEGATIONS

A. Plaintiff Is a Producer of Regional and Local News.

30. Plaintiff is among the most trusted sources of news in the communities it serves. As a producer of local news, Plaintiff occupies a unique and essential role in American civic life. Unlike national outlets, it covers school board meetings, municipal elections, community events, and other local issues that directly shape people's daily lives. Readers rely on it because it is close to the ground, accountable to its neighbors, and embedded in its community. Because of this constant presence in its communities, Plaintiff has consistently topped surveys as the nation's most trusted news source.

31. Plaintiff Boone is the publisher of several print and digital outlets focused on communities in Alabama, Georgia, Michigan, Minnesota, Mississippi, Virginia, and North Carolina.

32. Plaintiff expends significant time, money, and editorial effort in delivering quality reporting to its communities.

33. Plaintiff sustains these costly operations through various combinations of advertising sales, reader subscriptions, and content licensing. As set forth below, Defendants have

[3] *Id.*

[4] *Id.*

robbed Plaintiff of all three. By covertly scraping, copying, adapting, and using Plaintiff's content to train and deploy their GenAI products, Defendants have diverted traffic from Plaintiff's news sites, thereby causing a reduction in Plaintiff's advertising and subscription revenue while denying Plaintiff the licensing fees the Defendants should have been required to pay to use Plaintiff's protected works.

B. OpenAI and Microsoft Collaborated to Develop Transformational AI Technology.

1. OpenAI's Early History

34. OpenAI was incorporated in December 2015 as a nonprofit artificial intelligence research organization. Its founders, including Elon Musk, Reid Hoffman, Sam Altman, and Greg Brockman, launched the organization with a $1 billion investment toward the stated mission of "advancing digital intelligence" to "benefit humanity" while remaining "unconstrained by a need to generate a financial return."[5] OpenAI made this commitment publicly and repeatedly, representing to regulators, the public, and the research community that it would share its work openly and transparently with the world.

35. Those representations proved short-lived. Within three years, OpenAI began unwinding its nonprofit structure and constructing a fully commercial enterprise in its place. Today, OpenAI generates $2 billion in monthly revenue and was most recently valued at $852 billion.[6] In June 2026, OpenAI filed confidentially with the U.S. Securities and Exchange Commission for an initial public offering that could rank among the largest in the history of public markets.[7]

---

[5] *See* OpenAI, *Introducing OpenAI* (Dec. 11, 2015), https://openai.com/index/introducing-openai.

[6] OpenAI, *OpenAI raises $122 billion to accelerate the next phase of AI* (Mar. 31, 2026), https://openai.com/index/accelerating-the-next-phase-ai.

[7] Cade Metz, *OpenAI Files to Go Public as A.I. Companies Rush to Wall St.*, N.Y. Times (Jun. 8, 2026),

36. In 2022, OpenAI released ChatGPT, its flagship product. ChatGPT is an LLM-powered GenAI assistant chatbot. Users input prompts into the chatbot, and it delivers responses in plain language. GenAI systems like ChatGPT are able to produce plain language content because it learns from existing data—in this case, enormous quantities of human-authored written material.

37. ChatGPT became one of the fastest-adopted consumer products in history, reaching one million users within a month of its release and 100 million users within three months. Its growth has only continued since. Now, ChatGPT has over "900 million weekly active users, and over 50 million subscribers."[8] More than one year ago, OpenAI CEO Sam Altman described ChatGPT's reach as encompassing something on the order of 10% of the world's population.[9]

38. OpenAI offers a robust and expanding suite of LLM-powered services to its massive user base. Currently, OpenAI offers free and paid access to ChatGPT for individual consumers. The free version of ChatGPT provides users with limited access to GPT-5.5, OpenAI's latest model. Consumers can then pay a monthly fee, starting at $8-per-month and increasing to $100-per-month, for increasing access to GPT-5.5 and other features.

39. OpenAI has used similar subscription models throughout the relevant time period. For example, OpenAI previously made available a version of ChatGPT powered by GPT-3.5 to users for free, and OpenAI also offered a premium service, powered by OpenAI's then "most capable model" GPT-4, to consumers for $20 per month.

---

https://www.nytimes.com/2026/06/08/technology/openai-ipo.html.

[8] OpenAI, *OpenAI raises $122 billion to accelerate the next phase of AI* (Mar. 31, 2026), https://openai.com/index/accelerating-the-next-phase-ai.

[9] Beatrice Nolan, *Sam Altman Says '10% of the World Now Uses Our Systems a Lot' as Studio Ghibli-Style AI Images Help Boost OpenAI Signups*, Fortune (Apr. 14, 2025), https://fortune.com/2025/04/14/sam-altman-openai-user-base-doubled-few-weeks-10-of-world-uses-system.

40. OpenAI also offers business-focused subscriptions. These offerings include ChatGPT Enterprise and ChatGPT API tools designed to enable developers to incorporate ChatGPT into bespoke applications. OpenAI offers these tools to businesses via different payment options. For example, businesses may pay as they go to obtain OpenAI's "Business Codex," pay a monthly rate of $20 per user to obtain Business ChatGPT & Codex, or use OpenAI's custom pricing model to obtain ChatGPT Enterprise.

41. Upon information and belief, OpenAI plans to introduce new products that use similar methods and provide similar output (including Plaintiff's content).

42. OpenAI's models and products have been immensely valuable for OpenAI. As of 2026, public reporting estimated that over 92% of Fortune 500 companies are using ChatGPT.[10] As noted above, OpenAI generates $2 billion in revenue per month and has a post-money valuation of $852 billion.[11] Some analysts suggest OpenAI's IPO valuation could exceed $1 trillion, a number that exceeds the 2025 gross domestic product of all but 21 countries.[12]

43. Upon information and belief, all the OpenAI Defendants have been either directly involved in or have directed, controlled, and benefitted from OpenAI's widespread infringement and commercial exploitation of Plaintiff's protected works. OpenAI, Inc., alongside Microsoft, controlled and directed the widespread reproduction, distribution, and commercial use of Plaintiff's material perpetrated by OpenAI LP and OpenAI Global, LLC, through a series of holding and shell companies that include OpenAI Holdings, LLC, OpenAI GP, LLC, and OAI

[10] Sameer Khan, *50 ChatGPT Statistics Every Business Leader Should Know in 2026*, AI Business Weekly (Feb. 28, 2026), https://aibusinessweekly.net/p/chatgpt-statistics.

[11] *See* OpenAI, *OpenAI raises $122 billion to accelerate the next phase of AI* (Mar. 31, 2026), https://openai.com/index/accelerating-the-next-phase-ai.

[12] Douglas A. McIntyre, *OpenAI's $1 Trillion IPO*, Yahoo! Finance (Oct. 30, 2025), https://finance.yahoo.com/news/openai-1-trillion-ipo-141519224.html.

Corporation, LLC. OpenAI LP and OpenAI Global, LLC were directly involved in the design, development, and commercialization of OpenAI's GPT-based products, and directly engaged in the widespread stripping, copying, reproduction, distribution, and commercial use of Plaintiff's protected works. OpenAI LP and OpenAI Global, LLC also controlled and directed OpenAI, LLC and OpenAI OpCo, LLC, which were involved in distributing, selling, and licensing OpenAI's GPT-based products, and thus monetized the stripping, copying, reproduction, distribution, and commercial use of Plaintiff's works. As noted above, as of October 28, 2025, these entities are now under the organizational umbrellas of OpenAI Foundation and OpenAI Group PBC.

2. Microsoft's Role as Co-Architect

44. Microsoft has been an indispensable partner in virtually every aspect of OpenAI's commercial enterprise. In 2019, Microsoft made an initial $1 billion investment in OpenAI, the first installment of what would become a series of investments totaling approximately $13 billion. More than making passive investments, however, Microsoft became deeply intertwined with OpenAI as Microsoft became the primary architect of the physical and computational infrastructure on which OpenAI's models were trained and on which Plaintiff's copyrighted works were reproduced and misused at a massive scale. As Microsoft CEO Satya Nadella has said, Microsoft works on every aspect of the joint enterprise—performing optimizations, building tooling, and constructing infrastructure—and affirming the public view that the venture is a "joint project between Microsoft and OpenAI."

45. First, Microsoft and OpenAI jointly developed Azure-based AI supercomputing infrastructure, including a dedicated supercomputer that Microsoft built and operated for OpenAI's exclusive use. The system was designed to allow OpenAI to train its AI models on essentially the entire public internet. The system provided an integrated architecture comprising more than 285,000 Central Processing Unit ("CPU") cores, 10,000 Graphics Processing Units ("GPUs"), and

400 gigabits per second of network connectivity per GPU server, placing it among the five most powerful publicly known supercomputing systems in the world.[13] According to Nadella, Microsoft redesigned its data center infrastructure to be able to process OpenAI's massive training workloads, which were unlike any that had previously existed, and in order to provide the infrastructure that made OpenAI's frontier models achievable.[14]

46. In short, Microsoft "build[s] the infrastructure to train [OpenAI's] models. They're innovating on the algorithms and the training of these frontier models."[15]

47. Second, Microsoft has helped bring OpenAI's products to market by integrating them into Microsoft's own products. In February 2023, Microsoft launched Bing Chat, now rebranded as Copilot, a GenAI product powered by OpenAI's latest GPT model and integrated into Microsoft's search engine. Shortly thereafter, Microsoft and OpenAI jointly released Browse with Bing, a plug-in on the Bing browser allowing OpenAI's LLM to retrieve and summarize live web content.

48. Upon information and belief, like ChatGPT, these Microsoft products also provide responses to user input by taking content from Plaintiff's websites and delivering them to the user without necessarily directing the user to that site or including any attribution in its response.

49. As described in further detail below, by substituting their own product outputs for conventional search results, Defendants capture user engagement that would otherwise accrue to

---

[13] Jennifer Langston, *Microsoft announces new supercomputer, lays out vision for future AI work*, Microsoft (May 19, 2020), https://news.microsoft.com/source/features/ai/openai-azure-supercomputer/.

[14] *First on CNBC: CNBC Transcript: Microsoft CEO Satya Nadella Speaks with CNBC's Jon Fortt on "Power Lunch" Today*, CNBC (Feb. 7, 2023), https://www.cnbc.com/2023/02/07/first-on-cnbc-cnbc-transcript-microsoft-ceo-satya-nadella-speaks-with-cnbcs-jon-fortt-on-power-lunch-today.html.

[15] Microsoft, *Full Keynote: Satya Nadella at Microsoft Inspire 2023*, at 24:00, YouTube (Jul. 18, 2023), https://www.youtube.com/watch?v=RhwVMt_XCUE.

Plaintiff, depriving it of the user traffic and associated revenue its content would otherwise generate.

C. Defendants Improperly Used Plaintiff's Protected Works in Developing and Deploying their GenAI Products.

50. Defendants' massively lucrative enterprise relied on rampant copyright infringement. Microsoft and OpenAI created and distributed reproductions of Plaintiff's works while using those works to train their LLMs and deploying the GenAI products that incorporate them.

51. Microsoft and OpenAI improperly scraped, copied, reproduced, and used Plaintiff's protected works without authorization, both in training their LLMs and in deploying their GenAI products.

52. As described in further detail below, OpenAI's secrecy effectively made it impossible for copyright holders to determine whether their works had been swept into OpenAI's training pipeline without permission or compensation. It was not until years later that Plaintiff and other copyright holders would discover that they had.

53. As it pivoted from a nonprofit mission to profit-making, OpenAI also retreated from transparency—a move that has made it nearly impossible for copyright holders such as Plaintiff to uncover infringing activity. In its early years, OpenAI released its first two GPT models, GPT-1 and GPT-2, under open-source licenses and published substantial information about the resources it used to develop them. Beginning with its GPT-3 release in May 2020, however, OpenAI stopped disclosing meaningful information to the public about how its products were developed and what datasets it was using to train them. In fact, when it released GPT-4 in March 2023, OpenAI released a "technical report" in which it explicitly acknowledged that it was not disclosing "further details about the architecture (including model size), hardware, training

compute, dataset construction, training method, or similar."[16] When OpenAI released GPT-5 in 2025, it did not even release an accompanying "technical report." Instead, it produced a "System Card" that focuses on safety evaluations and results but contains very little, if any, details on architectural specifications or training data sources.[17]

D. Defendants' LLM Development Process.

54. Defendants' LLM development process starts with a collection stage, during which Defendants assemble and store massive training datasets drawn from text scraped across the internet, including from news and media sites like those of Plaintiff, reference sources, books, and other written works.

55. Defendants have used a series of such datasets across OpenAI's GPT model generations, including filtered versions of Common Crawl—a broad web archive encompassing billions of web pages dating back to 2008—as well as its own curated datasets, WebText and WebText2, which it assembled by curating high-quality sources of online content, such as news sites.

56. The "training" process involves repeatedly feeding these massive datasets, which inevitably include copyrighted content, through the LLM in an iterative process. Text from the training data is broken into "tokens," which are encoded units of words and sub-words, and stored in computer memory. The LLM is then programmed to repeatedly attempt to predict masked or missing tokens within sequences of text, adjusting billions of internal numerical parameters with each iteration to improve its predictions. Repeated with great frequency, this process results in an

[16] OpenAI, *GPT-4 Technical Report* (Mar. 22, 2023), https://cdn.openai.com/papers/gpt-4.pdf.

[17] OpenAI, *GPT-5 System Card* (Aug. 7, 2025), https://openai.com/index/gpt-5-system-card; *see also* Eric Landau, *GPT-5: A Technical Breakdown*, Encord (Aug. 8, 2025), https://encord.com/blog/gpt-5-a-technical-breakdown ("OpenAI has not released detailed architecture specs or training data sources.").

LLM whose parameters encode the statistical patterns—and, in many cases, the specific content—of the works on which it was trained. This process is what enables ChatGPT to generate responses to user prompts.

57. After initial training, LLMs are typically refined through fine-tuning, which involves additional training rounds using targeted content or human feedback, further shaping the model's outputs and deepening its encoding of specific material.

58. Researchers, and now copyright owners and courts, have observed that LLMs trained this way exhibit a behavior called "memorization."[18] Given the right prompt, LLMs will repeat large portions of materials on which they were trained. This phenomenon shows that LLM parameters encode retrievable copies of many of those training works.

59. After an LLM is trained, OpenAI's models are enhanced through a technique called retrieval-augmented generation, or "RAG," which takes place at the point of response generation rather than during training. Rather than relying solely on knowledge encoded in the model's parameters, a RAG-enabled system queries external sources in real time, including by crawling or indexing Plaintiff's websites, and incorporates the retrieved content directly into its generated responses.

E. Defendants Copied and Reproduced Plaintiff's Works Without Authorization while Training OpenAI's LLMs.

60. As described above, Defendants developed the LLMs that power OpenAI's GPT products by training and fine-tuning the models on massive datasets of human-made text material. Without permission or compensation, Defendants copied Plaintiff's materials to populate certain

[18] Gerrit J.J. van den Burg & Christopher K.I. Williams, *On Memorization in Probabilistic Deep Generative Models*, 35th Conference on Neural Information Processing Systems (Dec. 29, 2021), https://arxiv.org/pdf/2106.03216.

of the datasets they used to train various versions of their LLMs.

61. According to OpenAI's own reports, GPT-2 included 1.5 billion parameters and was developed using a training dataset OpenAI built called WebText, which includes "the text contents of 45 million links posted by users of the 'Reddit' social network." This includes the text contents of works owned by Plaintiff.

62. GPT-3, for its part, includes 175 billion parameters and was trained on a weighted combination of datasets, including the Common Crawl, WebText2, Books1, and Books2 datasets, as well as a dataset comprised of Wikipedia content.[19]

63. The most highly weighted dataset in GPT-3, Common Crawl, is a "copy of the Internet" made available by a 501(c)(3) organization of the same name that systematically scrapes the open internet, archiving petabytes (one petabyte is roughly one quadrillion bytes) of raw web pages, metadata, and text to make massive-scale data accessible to researchers, AI developers, and the public.[20]

64. A straightforward analysis of the Common Crawl dataset clearly shows that tens of thousands of Plaintiff's copyrighted articles and tens of millions of tokens worth of Plaintiff's content likely appeared in the Common Crawl training dataset Defendants used to train GPT-3.

65. The number of parameters OpenAI uses to train its GPT models has increased exponentially with each new version release. On information and belief, the Defendants have used, and continue to use, WebText, WebText2, Common Crawl, and other training datasets containing Plaintiff's protected content to train the GPT models.

---

[19] *Id.* at 9-10.

[20] *Commoncrawl Foundation*, ProPublica https://projects.propublica.org/nonprofits/organizations/261635908/202403189349101980/full; Common Crawl, *Frequently Asked Questions*, https://commoncrawl.org/faq.

66. Both OpenAI and Microsoft performed the infringing activity and collaborated in the unauthorized scraping, copying, storing, and reproduction of Plaintiff's works that were required to train the GPT models. Both OpenAI and Microsoft caused tens of thousands of Plaintiff's works to be copied and repeatedly run through the models.

67. As described above, Microsoft developed the underlying computing infrastructure for OpenAI so that the latter could scrape, copy, store, and reproduce the LLM training data at issue. Upon information and belief, Microsoft was intimately involved in every step of OpenAI's LLM training, from designing the models themselves to building the training datasets and supervising the models' training and fine-tuning. Regardless of whether Microsoft directly selected Plaintiff's works to train OpenAI's models, it knew or was willfully blind to the source of the data and that it was enabling and executing training processes using copyrighted material. Microsoft was able to control and prevent the infringing activity but failed to do so. And it reaped significant rewards as a result—both from its investment in OpenAI and in the incorporation of ChatGPT in its own products.

F. Defendants Reproduced and Publicly Displayed Plaintiff's Copyrighted Works Without Authorization while Deploying its GPT Products.

68. Defendants' infringement has also been observed in the outputs of their GPT Products. Researchers, courts, and other copyright holders have taken notice that various models of ChatGPT have demonstrated their underlying LLMs' propensity to "memorize" content on which they were trained and reproduce copies of it near-verbatim in response to targeted user prompts.

69. This phenomenon, well-documented in academic literature and confirmed by memorialized outputs from OpenAI's products, constitutes direct evidence that the works that appear in Common Crawl and other training datasets were copied during the training process and

that retrievable copies of their expressive content remain encoded within the models' parameters.

70. Plaintiff has been among the most affected copyright owners whose protected material is contained in the Defendants' training data. Investigative testing conducted in connection with substantially identical copyright infringement actions brought against the Defendants by The New York Times, the New York Daily News, the Chicago Tribune, the Denver Post, the Orlando Sentinel, the Sun Sentinel, the San Jose Mercury News, the Orange County Register, the St. Paul Pioneer Press, the Center for Investigative Reporting, The Intercept, Raw Story, AlterNet, and Ziff Davis ("News Plaintiff"), among others, has demonstrated that GPT-based models will reproduce substantial portions of published news articles when prompted appropriately, in some instances generating hundreds of words of verbatim text from a single article in a single response.

71. Across these cases, these News Plaintiff have supported their claims with side-by-side comparisons showing near-identical matches between AI-generated output and original published articles, establishing that News Plaintiff' content from Defendants' training data has been encoded into OpenAI's models through the LLM training process.

72. Upon information and belief, since these News Plaintiff conducted their investigative testing, OpenAI has disabled public access to the versions of ChatGPT that produced the demonstrably infringing, near-verbatim outputs they presented in their pleadings.

73. Upon information and belief, given that Defendants' training datasets, including WebText and Common Crawl, contained an enormous quantity of Plaintiff's protected content, the previous, now-offline versions of ChatGPT that these News Plaintiff tested would likewise generate near-verbatim copies of Plaintiff's copyrighted works if prompted to do so.

74. Each instance of a memorized news article constitutes an unauthorized copy or

derivative work belonging to that article's publisher that was used to train one or more of OpenAI's LLM models. Defendants therefore directly engaged in the unauthorized reproduction and publication of news articles in its training data to produce outputs from its GenAI products.

75. In addition to memorizing training data, Defendants' products can also generate infringing output through RAG functionality. As described above, RAG-enabled products such as ChatGPT Search and Deep Research retrieve content from Plaintiff's websites in real time and incorporate that content, oftentimes verbatim or in close paraphrase, into synthesized responses delivered to users. Unlike memorization, which reflects content encoded during training, RAG-based infringement occurs after a prompt from the user, at which point the model actively copies and processes Plaintiff's current content to produce its output.

76. As with their memorization testing, several News Plaintiff that have brought substantially identical copyright lawsuits against the Defendants have demonstrated that previous LLM models would, when prompted accordingly, retrieve and regurgitate protected news content that is publicly available on the internet.

77. Upon information and belief, in response to claims brought by News Plaintiff, OpenAI has disabled public access to the models of ChatGPT that produced the demonstrably infringing, near-verbatim outputs that the News Plaintiff presented in their pleadings. But prior to that, Defendants' LLM models when prompted accordingly would retrieve and regurgitate protected news content from Plaintiff's websites.

78. The presence of Plaintiff's copyrighted works in the Defendants' training data, combined with the copious amount of evidence that such training data has been regurgitated by the Defendants' products strongly suggests that Plaintiff's content can and has been reproduced and displayed publicly as a result of user prompts. Further information regarding the regurgitation

and output of Plaintiff's content is solely in OpenAI's possession, custody, and control.

G. Defendants' Infringement Was Willful.

79. Defendants' actions described above were performed willfully.

80. As two of the world's leading technology companies, Defendants possessed full knowledge of how their training models and products worked. All Defendants were involved in each step of the training, fine-tuning, testing, and commercialization of the GPT models.

81. OpenAI has expressly stated, "it would be impossible to train today's leading AI models without using copyrighted materials."[21] And Microsoft, through its statements and investment activity, has made clear that it is joined at the hip with OpenAI.

82. Defendants knew or should have known that the actions involved in scraping and copying of the Publisher's works on a massive scale was unauthorized. They also knew or should have known that the retention and encoding of these works within Defendants' models and systems was also unauthorized. Defendants also knew that these actions would result in the unauthorized display of such works that the models had either memorized or would present to users in the form of synthetic search results.

83. In fact, reports about the corporate infighting at OpenAI shows that in late 2023, Sam Altman reportedly clashed with OpenAI board member Helen Toner over a paper that Toner wrote criticizing the company over "safety and ethics issues related to the launches of ChatGPT and GPT-4, including regarding copyright issues."[22] OpenAI ousted, but subsequently reinstated at Microsoft's behest, Altman as CEO.

---

[21] OpenAI, *Written Evidence before the United Kingdom House of Lords Communications and Digital Select Committee inquiry: Large language models* (Dec. 5, 2023), at 4, https://committees.parliament.uk/writtenevidence/126981/pdf.

[22] Andrew Imbrie, Owen J. Daniels & Helen Toner, *Decoding Intentions*, Ctr. for Sec. & Emerging Tech. (Oct. 2023), at 29, https://cset.georgetown.edu/wp-content/uploads/CSET-Decoding-Intentions.pdf.

84. Plaintiff put Defendants on notice that these uses of Plaintiff's works were not authorized by placing copyright notices and linking to its terms of service (which contain, among other things, terms and conditions for the use of its works) on every page of its websites whose contents Defendants copied and displayed.

85. Upon information and belief, Defendants were aware of many examples of copyright infringement after ChatGPT, Browse with Bing, and Copilot (formerly Bing Chat) were released, some of which were widely publicized. These include multiple lawsuits dating back to 2023 and pending in this Court that allege such copyright infringement.

H. OpenAI Removed Copyright Management Information from Plaintiff's Works.

86. In 1998, Congress passed the Digital Millennium Copyright Act ("DMCA") in the wake of emerging technologies that could be used to evade existing copyright protections. Unlike copyright infringement claims, which require copyright owners to pay to register their works before bringing statutory claims, a DMCA claim does not require registration.

87. Among other things, the DMCA prohibits the knowing removal or alteration of CMI and the dissemination of works knowing that CMI has been removed or altered. *See* 17 U.S.C. § 1202(b). In so doing, the DMCA provided additional, broad protections to copyright owners such as Plaintiff—particularly as digital technology was evolving.

88. The DMCA defines the term "copyright management information," or CMI, as encompassing certain "information conveyed in connection with copies . . . of a work . . . , including in digital form," such as "the title and other information identifying the work, including the information set forth on a notice of copyright," "[t]he name of, and other identifying information about, the author of a work," "[t]he name of, and other identifying information about, the copyright owner of the work, including the information set forth in a notice of copyright,"

"[t]erms and conditions for use of the work," "[i]dentifying numbers or symbols referring to such information or links to such information," and "[s]uch other information as the Register of Copyrights may prescribe by regulation, except that the Register of Copyrights may not require the provision of any information concerning the user of a copyrighted work." 17 U.S.C. § 1202(c).

89. In its works, Plaintiff conveys CMI, such as Plaintiff's name, the titles, the authors' names, and terms and conditions for the use of the work.

90. Plaintiff also conveys its terms and conditions and copyright notice on the pages containing articles and other works.

91. Each Publisher conveys relevantly similar CMI content on its websites. No Publisher has licensed or otherwise permitted Defendants to include any of Plaintiff's works in their training datasets.

92. In compiling the datasets to train their products, Defendants not only scraped works from Plaintiff, but also intentionally removed Plaintiff's CMI from Plaintiff's works that were collected.

93. OpenAI has stated that it eschewed training language models on "a single domain of text" in favor of "building as large and diverse a dataset as possible."[23]

94. To do so, OpenAI "scraped all outbound links from Reddit, a social media platform, which received at least 3 karma" (a form of endorsement by Reddit users) to create a dataset that contained the text subset of 45 million links. That dataset was called WebText. As described above, an analysis performed by a technologist consultant employed by Plaintiff's counsel found that an open-source approximation of the WebText dataset called OpenWebText contained thousands of

[23] Alec Radford, *et al.*, *Language Models are Unsupervised Multitask Learners*, at 3, https://cdn.openai.com/better-languagemodels/language_models_are_unsupervised_multitask_learners.pdf.

records totaling millions of tokens of text from Plaintiff's websites.

95. OpenAI explained, "[t]o extract the text from HTML responses we use a combination of the Dragnet and Newspaper content extractors."[24]

96. Upon information and belief, OpenAI has continued to use the same or similar Dragnet and Newspaper text extraction methods when creating training sets for every version of ChatGPT since GPT-2, including Common Crawl. After publicly describing these methodologies for GPT-2, OpenAI has never indicated a change in their text extraction methodology for later models. Nor has OpenAI ever claimed to use any other text extraction methods for later versions of ChatGPT.

97. Upon information and belief, OpenAI used the Dragnet and Newspaper content extractors to intentionally remove CMI from Plaintiff's works, including author, title, copyright notice, and terms of use information.

98. For example, in an early article describing Dragnet, its creators describe the characteristics of the "content extractor." Specifically, the article explains, "[t]he goal of content extraction or boilerplate detection is to separate the main content from navigation chrome, advertising blocks, copyright notices and the like in web pages."[25] In a later update to "extend the original Dragnet algorithm," the design and purpose of Dragnet is described as to extract "chrome" containing certain CMI like copyright notices and footers, from the "main article content"—*i.e.*, the body of the work.[26]

---

[24] *Id.* (citations omitted).

[25] Matthew E. Peters & Dan Lecocq, *Content Extraction Using Diverse Feature Sets* (May 2013), https://archives.iw3c2.org/www2013/companion/p89.pdf.

[26] Matthew Peters, *Benchmarking Python Content Extraction Algorithms: Dragnet, Readability, Goose, and Eatiht* (Jan. 29, 2015), https://matt-peters.github.io/benchmarking-python-content-extraction-algorithms-dragnet-readability-goose-and-eatiht.

99. Upon information and belief, Dragnet is designed in such a way to separate out CMI that exists in the form of author, title, copyright notices, and term of use information from main article text—the ultimate goal of the extraction. So, if CMI exists outside the main article text, it is not extracted by Dragnet.

100. Newspaper is another content extractor that OpenAI used. One key feature of this content extractor is "avoiding IP blocks" on website. Upon information and belief, Newspaper is designed to separate and extract the article text on Plaintiff's webpages while removing Plaintiff's CMI, such as Plaintiff's terms and conditions and copyright notices.

101. While users of the Newspaper content extractor can choose to extract author and title information (but not copyright notices and term of use information), upon information and belief, OpenAI did not choose to do so. That is because Dragnet did not generally extract such information, and one purpose of using both datasets was to have consistent sets of data.

102. Upon information and belief, the reason that OpenAI chose to use Dragnet and Newspaper was because of these particular content extraction qualities. At the time OpenAI chose to use these content extractors, publicly available material on both extractors indicated that these methods remove author, title, copyright notices, and terms of use information from text. Upon information and belief, OpenAI intentionally and knowingly removed this copyright management information while assembling WebText.

103. Indeed, the use of both methods gave OpenAI backup methodologies in case a particular method did not yield optimal results for OpenAI's purposes. And having both methods would allow OpenAI to have consistency in training its models.

104. Upon information and belief, OpenAI would save the extractions from Plaintiff's relevant webpage in order to have a backup of the content for future use.

105. Numerous examples, including from other similarly situated parties who have brought claims, exist within OpenAI's WebText training dataset where a copy of the full text of a certain article is retained, but without CMI such as author, title, copyright notice, and term of use information. Upon information and belief, the same is true for Plaintiff's content that exists within WebText.

106. Upon information and belief, OpenAI employed the same or a similar CMI-stripping process when constructing the Common Crawl dataset it used to train its more advanced GPT models. Indeed, the C4 data set, which is a snapshot of the Common Crawl dataset that OpenAI used, contains the full text of online news articles by Plaintiff devoid of the publication name, article title, subtitle, byline, date, copyright notice, and terms of use links that appear on Plaintiff's webpages.

107. The lack of CMI from copies generated and retained in OpenAI's training datasets like WebText and Common Crawl, combined with fact that the extraction methods that OpenAI used were known to OpenAI, demonstrates that OpenAI acted intentionally in removing author, title, copyright notice, and terms of use information from Plaintiff's works. That harmed Plaintiff, which relies on CMI information to protect its copyright interests in its works.

108. Because OpenAI knew that its models were trained on datasets that had removed CMI, it knew that its chatbot products would likely disseminate unauthorized copies of Plaintiff's works without CMI to its customers.

109. OpenAI knew that Plaintiff's works would become outputs of OpenAI's GenAI products, conveyed without accompanying CMI and without permission. That harmed Plaintiff, which relies on CMI information to protect its copyright interests in its works.

110. The stripping of CMI from Plaintiff's works benefitted OpenAI—and upon

information and belief, OpenAI intended to receive that benefit. For one, had CMI not been removed, OpenAI's products would be trained on associating CMI with the text from Plaintiff's work. They would be more likely to deliver outputs that contain CMI, such as copyright and term of use information. Such a characteristic would create hesitation among users who, upon seeing that CMI associated with the outputs, would become concerned about unlawful use. For another, removing CMI in the training dataset would conceal Defendants' infringing activity because the outputs dissociated from CMI would not tip off the public, and importantly Plaintiff, that Defendants had illegally removed CMI in prior steps.

111. Moreover, OpenAI expressly contemplated the possibility that their customers "use or distribution of Output" may "infringe[] a third party's intellectual property right." In its Service Terms, OpenAI indemnifies its API customers for such infringements.[27] The provision exempts certain scenarios where the customer or customer end user took other actions against OpenAI policy. The commitment to paying costs and indemnifying against liability—all while it refuses to compensate those whose works it has pirated—demonstrates that OpenAI understood that even if customers are using ChatGPT according to OpenAI's own terms, they may still be engaging in infringing activity.

112. Thus, OpenAI knew that its removal of Plaintiff's CMI in this manner would induce, enable, conceal, or facilitate infringement by end-users of the GenAI products.

I. Defendants' Illegal Conduct Enriched Themselves at the Expense of Plaintiff.

113. Plaintiff has suffered and will continue to suffer concrete harm as a result of Defendants' unlawful conduct.

114. In creating and supporting local news, Plaintiff expends untold hours and many

[27] OpenAI, *Service terms*(last updated Jun. 12, 2026), https://openai.com/policies/service-terms.

millions of dollars to develop news-gathering, storytelling, community engagement, and information-conveying capabilities. Such activities are not only private, business interests: they also deliver a crucial public good in our democratic society.

115. Plaintiff's economic viability to undertake these actions depends on readers who purchase and renew subscriptions to Plaintiff's products or otherwise engage with Plaintiff's content and advertising. That includes accessing Plaintiff's print publications, websites—with and without paywalls—mobile applications, newsletters, e-editions, and other services. Plaintiff also licenses or otherwise agrees to provide its content to other entities, such as other media entities, museums, and libraries under contractual agreements. When it does so, Plaintiff receives proper renumeration for the use of its works, and such agreements come with clear parameters around the display and use of the works.

116. Whenever prospective readers access Plaintiff's works through alternate means, without any compensation to Plaintiff, that harms Plaintiff's interests in several ways.

117. First, it directly diminishes revenue from subscriptions, licensing agreements, and other direct contracts by siphoning readers.

118. Second, it diminishes advertising revenue by reducing the readership base when readers are less likely to read print periodicals or visit Plaintiff's websites and other publication displays.

119. Third, it disincentivizes third parties such as licensers and advertisers from entering into agreements with Plaintiff.

120. Fourth, it damages Plaintiff's ability to attract and retain journalism talent, creating a cyclical effect on Plaintiff's business and further eroding the local media industry upon which this nation relies.

121. The above harms to Plaintiff come at an acute time in the journalism industry. Local journalism operates in a fragile economic environment. But its impact on our civil society and our democracy cannot be overstated. As a Democracy Fund analysis of numerous studies demonstrates, local journalism has demonstrably increased civic participation, produced greater cohesiveness in communities, and reduced public corruption.[28]

122. Moreover, at a time when the American public is deeply divided, local news remains a beacon of hope and a uniquely trusted source of information. According to the Pew Research Center, 71% of Americans believe their local news outlets are doing a good job at reporting the news accurately. Sixty-two percent believe local news is effective at dealing fairly with all sides.[29]

123. Defendants' illegal and unauthorized use of Plaintiff's works as described above wreaks havoc on the viability of local journalism. By taking the Publisher's works without attribution or compensation, Defendants threaten Plaintiff's business model in fundamental ways.

124. By contrast, Defendants have reaped astronomical financial benefits off the backs of Plaintiff.

125. In March 2026, OpenAI announced that it has a post money valuation of $852 billion.[30] It touts $2 billion in revenue per month and over 900 million active users weekly.[31] Its clients include millions of individuals and thousands of companies, including nearly all Fortune

---

[28] Josh Stearns & Christine Schmidt, *How We Know Journalism is Good for Democracy*, Democracy Fund (Sept. 15, 2022), https://democracyfund.org/idea/how-we-know-journalism-is-good-for-democracy.

[29] *See* John Gramlich, *Q&A: What Pew Research Center's new survey says about local news in the U.S.*, Pew Research Ctr. (Mar. 16, 2019), https://www.pewresearch.org/short-reads/2019/03/26/qa-what-pew-research-centers-new-survey-says-about-local-news-in-the-u-s.

[30] *See* OpenAI, *OpenAI raises $122 billion to accelerate the next phase of AI* (Mar. 31, 2026), https://openai.com/index/accelerating-the-next-phase-ai.

[31] *See Id.*

500 companies, and many on paying subscriptions.[32]

126. In June 2026, OpenAI filed registration statements ahead of an IPO with the Securities and Exchange Commission.[33] Some analysts suggest OpenAI's IPO valuation could exceed $1 trillion, a number that exceeds the 2025 gross domestic product of all but 21 countries.[34]

127. Microsoft's investment in OpenAI has proven extremely lucrative. After its initial $1 billion investment in 2019, Microsoft reportedly added another $12 billion to its investment, rendering its stake in the business to be approximately 27%. As noted above, OpenAI is preparing one of the most lucrative IPOs in history, and Microsoft is its biggest shareholder.

128. Moreover, Microsoft's own products have stood to gain from the enterprise. Its Bing search engine's popularity rose after it integrated ChatGPT. And Microsoft also uses ChatGPT in its Copilot product, which is streamlined with its Office 365 premium products subscription. Microsoft's reported revenue in March 2026 was $82.9 billion, nearly a 20% increase year over year.[35]

129. Defendants' successes would not have been possible without the use of copyrighted material such as Plaintiff's works.

---

[32] *See* Sameer Khan, *50 ChatGPT Statistics Every Business Leader Should Know in 2026*, AI Business Weekly (Feb. 28, 2026), https://aibusinessweekly.net/p/chatgpt-statistics

[33] Adam Levy, *OpenAI Just Took the First Step Towards Its IPO. Here's How to Invest in the Stock Now.*, The Motley Fool (Jun. 14, 2026), https://www.fool.com/investing/2026/06/14/openai-just-took-the-first-step-toward-its-ipo-her.

[34] Douglas A. McIntyre, *OpenAI's $1 Trillion IPO*, Yahoo! Finance (Oct. 30, 2025), https://finance.yahoo.com/news/openai-1-trillion-ipo-141519224.html.

[35] Microsoft, *Microsoft Cloud and AI strength fuels third quarter results* (Apr. 29, 2026), https://news.microsoft.com/source/2026/04/29/microsoft-cloud-and-ai-strength-fuels-third-quarter-results.

CLAIMS FOR RELIEF

COUNT I

Copyright Infringement in Violation of 17 U.S.C. § 501

130. Plaintiff incorporates by reference the preceding factual allegations.

131. Plaintiff is the owner of registered copyrights.

132. Plaintiff has a business practice of publishing in electronic format on its websites articles that also appear in that newspaper's print edition. It has additionally been the business practice of Plaintiff to publish in electronic format on its websites articles from older print editions of those newspapers. The electronic versions of the articles are substantially the same as their print-edition counterparts.

133. As the owner of the registered copyrights in these works, Plaintiff holds the exclusive rights to the works under 17 U.S.C. § 106.

134. Electronic versions of the articles published by Plaintiff were copied to train Defendants' GPT models and, in many cases, have been distributed by and encoded within Defendants' GPT models.

135. By illegally building training datasets containing Plaintiff's works, including by scraping copies of Plaintiff's works from Plaintiff's websites and reproducing these works from third-party datasets, Microsoft and the OpenAI Defendants have directly infringed Plaintiff's exclusive rights in its copyrighted works.

136. By illegally storing, processing, and reproducing the training datasets containing Plaintiff's works to train the GPT models on Microsoft's supercomputing platform, Microsoft and the OpenAI Defendants have jointly directly infringed Plaintiff's exclusive rights in its copyrighted works.

137. On information and belief, by storing, processing, and reproducing the GPT models

trained on Plaintiff's works, which GPT models themselves have memorized, on Microsoft's supercomputing platform, Microsoft and the OpenAI Defendants have jointly directly infringed Plaintiff's exclusive rights in its copyrighted works.

138. By disseminating generative output containing copies and derivatives of Plaintiff's works through the ChatGPT offerings, the OpenAI Defendants have directly infringed Plaintiff's exclusive rights in its copyrighted works.

139. By disseminating generative output containing copies and derivatives of Plaintiff's works through the Copilot (formerly known as Bing Chat) offerings, Microsoft has directly infringed Plaintiff's exclusive rights in its copyrighted works.

140. On information and belief, Defendants' infringing conduct alleged herein was and continues to be willful and carried out with full knowledge of Plaintiff's rights in its works. As a direct result of their conduct, Defendants have wrongfully generated revenue from copyrighted works that they do not own.

141. By and through the actions alleged above, Defendants have infringed and will continue to infringe Plaintiff's copyrights.

142. As a direct and proximate result of Defendants' infringing conduct alleged herein, Plaintiff has sustained and will continue to sustain substantial, immediate, and irreparable injury for which there is no adequate remedy at law. Unless Defendants' infringing conduct is enjoined by this Court, Defendants have demonstrated an intent to continue to infringe Plaintiff's works. Plaintiff therefore is entitled to permanent injunctive relief restraining and enjoining Defendants' ongoing infringing conduct.

143. Plaintiff is further entitled to recover statutory damages, actual damages, restitution of profits, attorney's fees, and other remedies provided by law.

COUNT II

Vicarious Copyright Infringement

144. Plaintiff incorporates by reference the preceding factual allegations.

145. Microsoft controlled, directed, and profited from the infringement perpetrated by the OpenAI Defendants. Microsoft controls and directs the supercomputing platform used to store, process, and reproduce the training datasets containing the Plaintiff's works, the GPT models, and OpenAI's ChatGPT offerings. Microsoft had the ability to control the infringing activity perpetrated by the OpenAI defendants but failed to do so. Microsoft profited from the infringement perpetrated by the OpenAI Defendants by incorporating the infringing GPT models trained on the Plaintiff's works into its own product offerings, including Copilot (formerly known as Bing Chat).

146. Defendants OpenAI, Inc.; OpenAI, GP; OAI Corporation, LLC; OpenAI Holdings, LLC; and Microsoft controlled, directed, and benefitted from the infringement perpetrated by Defendants OpenAI LP; OpenAI Global, LLC; OpenAI OpCo, LLC; OpenAI, LLC; and OpenAI Global PBC including the reproduction and distribution of the Plaintiff's works.

147. Defendants OpenAI Global, LLC and OpenAI LP directed, controlled, and benefitted from the infringement perpetrated by Defendants OpenAI OpCo, LLC and OpenAI, LLC, including the reproduction and distribution of the Plaintiff's works.

148. Defendants OpenAI, Inc.; OpenAI LP; OAI Corporation, LLC; OpenAI Holdings, LLC; OpenAI Global, LLC; and Microsoft are vicariously liable for copyright infringement.

COUNT III

Violation of the Digital Millenium Copyright Act, 17 U.S.C. § 1202(b)

149. Plaintiff incorporates by reference the preceding factual allegations.

150. Plaintiff included one or more forms of CMI in each of Plaintiff's works, including: a copyright notice, authors' names, publisher's name, title and other identifying information, terms

and conditions of use, and identifying numbers or symbols referring to the copyright-management information.

151. Without Plaintiff's permission, OpenAI copied Plaintiff's works and used them as training data for their GenAI models.

152. On information and belief, OpenAI removed Plaintiff's CMI in building the training datasets containing copies of Plaintiff's works, including removing Plaintiff's CMI from Plaintiff's works scraped directly from Plaintiff's websites and removing Plaintiff's CMI from Plaintiff's works reproduced from third-party datasets.

153. On information and belief, OpenAI removed Plaintiff's CMI through the generation of synthetic search results, including removing Plaintiff's CMI when scraping Plaintiff's works from Plaintiff's websites and generating copies or derivatives of Plaintiff's works as the output of ChatGPT offerings.

154. On information and belief, OpenAI removed Plaintiff's CMI in generating outputs from the GPT models containing copies or derivatives of Plaintiff's works.

155. By design, OpenAI's GPT-based products do not preserve any CMI, and the outputs of Defendants' GPT models removed any copyright notices, titles, and identifying information, despite the fact that those outputs were often verbatim reproductions of Plaintiff's works. Therefore, OpenAI intentionally removed CMI from Plaintiff's works in violation of 17 U.S.C. § 1202(b)(1).

156. OpenAI's removal or alteration of Plaintiff's CMI was done knowingly and with the intent to induce, enable, facilitate, or conceal OpenAI's or end-users' infringement of Plaintiff's copyrights.

157. OpenAI knew or had reasonable grounds to know that their removal of CMI would

facilitate copyright infringement by concealing the fact that the GPT models are infringing copyrighted works and that outputs from the GPT models are infringing copies and derivative works.

158. Plaintiff has been injured by OpenAI's removal of CMI. Plaintiff is entitled to statutory damages, actual damages, restitution of profits, and other remedies provided by law, including full costs and attorney's fees.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff demands judgment against each Defendant as follows:

i. Awarding Plaintiff statutory damages, including willful infringement damages, compensatory damages, restitution, disgorgement, and any other relief that may be permitted by law or equity;

ii. Permanently enjoining Defendants, their agents and employees, and all person acting in concert or participation with Defendants, from the unlawful, unfair, and infringing conduct alleged herein;

iii. An injunction under 17 U.S.C. § 503(b) requiring Defendants to remove all copies of Registered Works from all GPT or other LLM models and training sets;

iv. An award of costs, expenses, and attorney's fees as permitted by law; and

v. Such other or further relief as the Court may deem appropriate, just, and equitable.

DEMAND FOR JURY TRIAL

Plaintiff hereby demands a jury trial for all claims so triable.

Dated: October 30, 2026.

Respectfully submitted,

/s/ Don Barrett
John W. ("Don") Barrett (MSB #2063)

Katherine Barrett Riley (MSB #99109)
Sterling Aldridge (MSB #104277)
David McMullan, Jr. (MSB #8494)
BARRETT LAW GROUP, P.A.
P.O. Box 927
404 Court Square North
Lexington, Mississippi 39095
(662) 834-2488
dbarrett@barrettlawgroup.com
kbriley@barrettlawgroup.com
saldridge@barrettlawgroup.com
dmcmullan@barrettlawgroup.com

Wilson H. Carroll
WILSON CARROLL, PLLC
2506 Cherry St.
Vicksburg, MS 39180
(601) 953-6579
wilson@wilsoncarroll.com

Charles J. LaDuca
Sam Pritchett
CUNEO GILBERT FLANNERY & LADUCA, LLP
2445 M Street, N.W., Ste. 740
Washington, D.C. 20037
(202) 789-3960
charlesl@cuneolaw.com
spritchett@cuneolaw.com

Charles Barrett
CUNEO GILBERT FLANNERY & LADUCA, LLP
4235 Hillsboro Pike, Ste. 100
Nashville, TN 37215
(202) 844-3647
cbarrett@cuneolaw.com

Christian Hudson
CUNEO GILBERT FLANNERY & LADUCA, LLP
222 Livingston Street
Brooklyn, NY 11201
(202) 789-3960
christian@cuneolaw.com
aklein@cuneolaw.com

Counsel for Plaintiff