UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

ZIFF DAVIS, INC., ZIFF DAVIS, LLC, IGN ENTERTAINMENT, INC., EVERYDAY HEALTH MEDIA, LLC, MASHABLE, INC., and CNET MEDIA, INC., Plaintiffs, v. OPENAI, INC., OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO, LLC, OPENAI GLOBAL LLC, OAI CORPORATION, and OPENAI HOLDINGS, LLC, Defendants.	No. 25-cv-04315-SHS-OTW No. 25-md-03143-SHS-OTW **SECOND AMENDED COMPLAINT** **JURY TRIAL DEMANDED**

SECOND AMENDED COMPLAINT

Plaintiffs Ziff Davis, Inc., Ziff Davis, LLC, IGN Entertainment, Inc., Everyday Health Media, LLC, Mashable, Inc., and CNET Media, Inc. (together, "**Ziff Davis**" or "**Plaintiffs**"), for their Second Amended Complaint against Defendants OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OpCo LLC, OpenAI Global LLC, OAI Corporation, and OpenAI Holdings, LLC (together, "**OpenAI**"), respectfully allege as follows:

INTRODUCTION

1. Ziff Davis, by its attorneys, brings this action against OpenAI for copyright infringement, violations of the Digital Millennium Copyright Act ("**DMCA**"), and unjust enrichment.

2. Ziff Davis has published high-quality journalism for nearly 100 years, growing from its roots as the publisher of *Popular Aviation* to its current stewardship of over 45 diverse digital media publications and internet brands, including *IGN*, *Mashable*, *CNET*, *ZDNET*, *PCMag*,

Lifehacker, *BabyCenter*, and *Everyday Health*. Each year, it produces nearly 2 million new articles and article updates—including over 5,000 product reviews—in which it owns the exclusive rights under 17 U.S.C. § 106 of the Copyright Act ("**Ziff Davis Works**" or "**Plaintiffs' Works**").

3. OpenAI develops Artificial Intelligence ("**AI**") large language models ("**LLMs**"), including its "Generative Pretrained Transformer" or "GPT" series of LLMs.

4. To build and operate its LLM software and LLM-based products and services, OpenAI has intentionally and relentlessly reproduced exact copies and created derivatives of Ziff Davis Works without Ziff Davis's authorization.

5. Specifically, Defendant OpenAI has and continues to knowingly:

a. Copy the text of Ziff Davis Works from Ziff Davis's websites without authorization;

b. Violate and circumvent Ziff Davis's explicit written terms and demands, and technological controls;

c. Strip out key copyright management information ("**CMI**") from Ziff Davis Works, and distribute copies of those Works without CMI;

d. Use copied Ziff Davis Works to develop LLMs;

e. Use copied Ziff Davis Works to operate LLMs and LLM-based products and services, which it provides to third parties;

f. Reproduce, distribute, display, perform, and make available for access, Ziff Davis Works verbatim and in close paraphrase and derivative form (but with CMI removed);

6. Defendant OpenAI has taken each of these steps knowing that they violate Ziff Davis's intellectual property rights and the law.

7. OpenAI's actions violate, either directly or indirectly:

a. Ziff Davis's exclusive rights of reproduction, preparation of derivative works, distribution, performance, and public display under the Copyright Act;

b. Provisions of the DMCA prohibiting circumvention of technical copy protection measures and removal of copyright management information;

c. Delaware state laws' protection against unjust enrichment; and

8. OpenAI's actions and violations of law harm Ziff Davis because they:

a. Substitute OpenAI's LLMs, products, services, and outputs for Ziff Davis's media content and distribution services;

b. Usurp Ziff Davis's ability to monetize user interactions through advertising, product sales commissions, and other revenue-producing activities;

c. Deprive Ziff Davis of the licensing fees OpenAI should have paid to Ziff Davis; and

9. Upon information and belief, OpenAI has long been aware that its use of third-party works, including Ziff Davis Works, constitutes copyright infringement, and further has been aware of the widespread reports of copyright infringement after releases of its various products. Further, by letters dated February 5, 2024 and May 20, 2024, Ziff Davis put OpenAI on notice of Ziff Davis's claims.

10. Ziff Davis also wrote to OpenAI on August 23, 2024 to provide further information that would support an infringement claim and to request a meeting to discuss a licensing arrangement. OpenAI rebuffed the invitation to meet.

11. Ziff Davis has documented—even pre-discovery—specific instances in which OpenAI copied Ziff Davis's content to develop and operate LLMs, violating the DMCA in multiple knowing and intentional ways in the process.

12. OpenAI has concealed much of its misconduct by abandoning its founding principle of openness—and constituent open-sourcing and transparent publication practices—deferring a wealth of information about OpenAI's infringements and violations of law for discovery.

13. Ziff Davis alleges and will prove that whole portions of OpenAI's LLMs are themselves infringing copies of Ziff Davis Works.

14. OpenAI seeks to move fast and break things on the assumption that the federal courts will not be able to effectively redress content owners' sometimes existential concerns before it is too late.

15. Simultaneously with intentionally and egregiously exploiting the content of Ziff Davis and other commercial web publishers without permission, OpenAI is also actively creating and cultivating a market to license content from publishers.

16. OpenAI, by its actions, has flouted copyright law and discredited its own pretext for that flouting. Ziff Davis therefore seeks relief from this Court.

JURISDICTION AND VENUE

17. This Court has subject matter jurisdiction over Ziff Davis's claims pursuant to 28 U.S.C. §§ 1331 and 1338 as they arise under the Copyright Act.

18. This Court has personal jurisdiction over each of the OpenAI Defendants because they are incorporated in, registered in or formed in Delaware, and therefore reside in this District.

Venue is proper in this District pursuant to 28 U.S.C. §§ 1391(b) and 28 U.S.C. § 1400(a) because OpenAI and its agents reside in or may be found in this District.

PARTIES

19. Plaintiff Ziff Davis, Inc. is a Delaware corporation with its principal place of business at 360 Park Avenue South, 17th Floor, New York, New York.

20. Plaintiff Ziff Davis, LLC is a Delaware limited liability company with its principal place of business at 360 Park Avenue South, 17th Floor, New York, New York.

21. Plaintiff IGN Entertainment, Inc. is a Delaware corporation with its principal place of business at 360 Park Avenue South, 17th Floor, New York, New York.

22. Plaintiff Everyday Health Media, LLC is a Delaware limited liability company with its principal place of business at 360 Park Avenue South, 17th Floor, New York, New York.

23. Plaintiff Mashable, Inc. is a Delaware corporation with its principal place of business at 360 Park Avenue South, 17th Floor, New York, New York.

24. Plaintiff CNET Media, Inc. is a Delaware corporation with its principal place of business at 360 Park Avenue South, 17th Floor, New York, New York.

25. Defendant OpenAI, Inc. is a Delaware corporation with its principal place of business at 3180 18th Street, San Francisco, California. Upon information and belief, Defendant OpenAI, Inc. is a non-profit company that governs other OpenAI entities and perpetrated the infringing and other unlawful activities alleged in this Second Amended Complaint.

26. Defendant OpenAI GP, LLC is a Delaware limited liability company with its principal place of business at 3180 18th Street, San Francisco, California. OpenAI GP, LLC is wholly owned and controlled by OpenAI, Inc. and is responsible for managing and operating the day-to-day business and affairs of OpenAI OpCo LLC, formerly known as OpenAI LP. Through

OpenAI GP, LLC, OpenAI, Inc. controls OpenAI OpCo LLC and OpenAI Global LLC. Upon information and belief, OpenAI, Inc. relaunched itself as a for-profit enterprise in 2019, specifically through OpenAI GP, LLC and OpenAI OpCo LLC, and these entities perpetrated the infringing and other unlawful conduct alleged in this Second Amended Complaint.

27. Defendant OpenAI, LLC is a Delaware limited liability company, with its principal place of business located at 3180 18th Street, San Francisco, California. OpenAI, LLC was formed in September 2020, is responsible for monetizing and distributing OpenAI's LLM-based products, and upon information and belief perpetrated the infringing and other unlawful activities alleged in this Second Amended Complaint. Upon information and belief, OpenAI, LLC is a subsidiary of OpenAI Global LLC, which is owned and controlled by both OpenAI, Inc. and Microsoft Corporation.

28. Defendant OpenAI OpCo LLC is a Delaware limited liability company with its principal place of business at 3180 18th Street, San Francisco, California. OpenAI OpCo LLC, formerly known as OpenAI LP, is a subsidiary of OpenAI Global LLC, and is the sole member of OpenAI, LLC. Upon information and belief, OpenAI OpCo LLC perpetrated the infringing and other unlawful activities alleged in this Second Amended Complaint, directing this activity through its control of OpenAI, LLC. Additionally, OpenAI OpCo LLC serves as the for-profit arm of OpenAI, overseeing the commercialization of OpenAI's products and services, including LLMs and associated application programming interfaces ("**APIs**").

29. Defendant OpenAI Global LLC is a capped profit limited liability company formed under the laws of Delaware in late 2022, with its principal place of business located at 3180 18th Street, San Francisco, California. Upon information and belief, OpenAI Global LLC is a subsidiary of Open AI, Inc. OpenAI, Inc. holds a majority interest in OpenAI Global LLC

indirectly through OpenAI GP, LLC. Upon information and belief, OpenAI Global LLC perpetrated the infringing and other unlawful activities alleged in this Second Amended Complaint through its ownership, control, and direction of OpenAI, LLC.

30. Defendant OAI Corporation is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, California. OAI Corporation is a subsidiary of OpenAI, Inc. and its sole member is OpenAI Holdings, LLC. Upon information and belief, OAI Corporation perpetrated the infringing and other unlawful activities alleged in this Second Amended Complaint through its ownership, control, and direction of OpenAI Global LLC and OpenAI, LLC.

31. Defendant OpenAI Holdings, LLC is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, California. Its sole members are OpenAI, Inc. and Aestas, LLC, the latter of which is wholly owned by Aestas Management Company, LLC. Aestas Management Company, LLC is a Delaware entity established to facilitate a $495 million capital raise for OpenAI. Upon information and belief, OpenAI Holdings, LLC perpetrated the infringing and other unlawful activities alleged in this Second Amended Complaint.

FACTUAL ALLEGATIONS

Ziff Davis's Media Business

32. Founded in 1927 by William Ziff and Bernard Davis, Ziff Davis has been a cornerstone of publishing and media innovation for nearly a century. From its origins in enthusiast print magazines to becoming a digital media powerhouse, Ziff Davis has always adhered to William Ziff's enduring principle that "growth will come in those media that reach the…consumer in an intensive way, not in those that touch everyone lightly." This philosophy continues to define

Ziff Davis's brands, driving their ability to deliver deeply relevant, expertly-researched, and trusted content.

33. Today, Ziff Davis is a vertically focused digital media and technology company with over 45 portfolio media brands. Its well-known publications and editorial sites, including *IGN*, *Mashable*, *CNET*, *ZDNET*, *PCMag*, *Lifehacker*, *BabyCenter*, and *Everyday Health*, are circulated to millions of consumers on a daily basis. A list of Ziff Davis brands is attached hereto as Exhibit A.

34. Ziff Davis's storied history and vast audience of readers demonstrate the quality and value of its brand and content.

Ziff Davis's Media Verticals

35. Ziff Davis provides authoritative content and services in discrete and consumer-rich media "vertical" categories, including Technology and Shopping, Gaming and Entertainment, and Health and Wellness, which operate under their own well-known umbrella brands:[1]

- **Technology and Shopping.** Ziff Davis's CNET Group comprises marquee names of technology journalism, including *CNET*, a longtime leader in providing expert news, information, and trusted reviews covering consumer technology, money, wellness, and other topics; *ZDNET*, an authoritative provider of global technology business news, advice, and insights; *Mashable*, a news and entertainment site covering technology and digital media; *PCMag*, a leading authority on technology that delivers lab-based, independent reviews of the latest products; *Lifehacker*, which provides users with general life tips and tech help; *AskMen*, which provides expert advice

[1] Ziff Davis also operates well-known connectivity, cybersecurity, and marketing technology businesses.

in men's lifestyle; and *Spiceworks*, a global IT professional network and marketplace that publishes news and research for technology professionals. Ziff Davis Shopping provides expert editorial content for shopping needs across eight brands including *RetailMeNot*, which helps guide consumers through the best shopping deals; *Deals of America*, which collates the best online deals, coupons, promotions and special offers from US stores in real time; *TechBargains*, the leading online destination for consumers searching for the best deals on the hottest products and gadgets; and *BlackFriday* and *TheBlackFriday*, which help shoppers view popular Black Friday advertisements in one central location.

- **Gaming and Entertainment.** Ziff Davis's IGN Entertainment is a leading digital media company in its own right and predominates in its sector in the quality and influence of its content. It provides extensive authoritative content relating to video games (including game help and guide content), films, anime, television, comics, and other media. It draws on unparalleled access to game developers and the broader gaming community. Its brands include *IGN*, *Gamer Network* (and constituent brands including *Eurogamer*, *Rock Paper Shotgun*, and *VG247*), *Maxroll*, and *MapGenie.* IGN Entertainment is considered one of the most recognizable brands for games, entertainment, and fan-culture in major markets worldwide.
- **Health and Wellness.** Ziff Davis's Everyday Health Group is a recognized leader in patient and provider news, information, and services. Its most recognizable assets include *Everyday Health*, a provider of trustworthy

health and wellness information through expert editorial staff and medical reviewers; *What To Expect*, the world's most recognized and trusted pregnancy and parenting brand; *BabyCenter*, on which parents and expecting parents rely for content about conception, pregnancy, birth, and early childhood development; *MedPage Today*, a trusted source for clinical news coverage across the medical specialties; and *theSkimm*, a provider of news and information through its website and suite of email newsletters.

36. Ziff Davis's media properties reach, inform, and guide the decision-making of hundreds of millions of consumers. Ziff Davis's top digital media properties together averaged over 291 million unique user visits per month in the last twelve available months of third-party traffic data.

37. Ziff Davis has built a legacy of trust and authority across diverse audiences, cementing its media properties as leaders in their respective fields. Its most widely recognized media properties have won numerous awards, including:

a. *CNET*, *Mashable*, and *Lifehacker* are frequently honored for the excellence of their digital media content, winning awards from the LA Press Club, FOLIO's *Eddie & Ozzie Awards*, *The Webby Awards,* and the *Telly Awards.*

b. *MedPage Today* is a frequent winner of the Health Information Resource Center's *Digital Health Awards*, recognizing the exceptional quality of its deep investigative reporting, news and opinion content, and video series. *MedPage Today* has won 41 *Digital Health Awards* in the past three awards years.

c. *What to Expect* and *BabyCenter* have received a variety of awards for their news, storytelling, and interactive media, including multiple *The Webby Awards* and *Digital Health Awards*.

d. *Everyday Health* has won over 100 awards in the past three years alone, including many *Digital Health Awards*, Academy of Interactive & Visual Arts *w3 Awards* for best website editorial experience, and *Medical Marketing & Media Awards* for best media brand.

e. *IGN*, *Eurogamer*, and *VG247* have earned dozens of awards throughout their decades of operation, including *MCV/DEVELOP Awards*, *Games Media Awards*, *Association of Online Publishers Digital Publishing Awards*, *Shorty Awards*, and *Webby Awards*.

38. Reporters for *Mashable*, *CNET*, *ZDNET*, *IGN*, *MedPage Today*, and other Ziff Davis investigative reporters devote themselves tirelessly to original investigation, research, and reporting, and they consistently and frequently break major stories in their respective fields.

39. Ziff Davis's content is supported by rigorous research, reporting, and product testing, and its experienced writers and editors incorporate the results of these efforts into original stories. For example:

a. Since 1984 *PCMag* has operated one of the longest-running independent testing facilities for consumer technology products in the country. It produces unbiased technology product and service reviews, and *PCMag*'s *Editor's Choice* award is recognized globally as a trusted mark for buyers and sellers of technology products and services.

b. *CNET* maintains three specialized lab facilities: one in New York City for testing of technology products and measurement of performance data such as battery life, screen brightness, color fidelity, and network connection lag, which it shares with *ZDNET*; a second in Louisville, Kentucky, in which products such as large appliances, televisions, smart home devices, and fitness equipment are tested under strict environmental controls; and a third in Reno, Nevada for mattress review testing.

c. *Everyday Health*, *What to Expect*, and *BabyCenter* all produce journalism that is extensively researched and reported by its original authors and rigorously fact-checked by teams of expert medical reviewers holding M.D.s, Ph.D.s, master's degrees, and other specialized degrees.

d. The journalists writing for *IGN* and Ziff Davis's other Gaming and Entertainment digital media properties spend hours playing and beating video games and critically consuming entertainment content, and they use this research and experience to create original reviews, reporting, and book-length game guides.

40. Ziff Davis invests directly in content creation through its teams of journalists, editors, designers, and producers, as well as indirectly, by funding website and app platform development, innovation, and maintenance. Ziff Davis invests in the corporate infrastructure that supports that content creation, including through the work of its human resources, finance, accounting, business and corporate development, technology, product management, information security, legal, and management professionals who work within its operating divisions and at the corporate level.

41. Ziff Davis has a workforce of over 3,800 employees and over 40 offices worldwide.

42. In short, Ziff Davis and its personnel expend and invest considerable time, energy, and resources—and take financial and journalistic risks—to create the Ziff Davis Works.

43. Ziff Davis's reach is global: its media properties offer approximately 70 regional editions, and its content is published in 20 languages in over 100 countries. Ziff Davis also has a robust social media presence and as of June 2025 boasted approximately 122 million followers across various platforms.

44. In the aggregate, Ziff Davis's websites realize more inbound traffic from traditional, link-oriented search engines than its key peers, for example receiving over 195 million visits directly from search engine results in March 2025 according to third-party reporting:

Rank	Publisher	March 2025 organic search visits
1	**Ziff Davis**	**195,444,230**
2	IAC	171,992,393
3	NY Times	167,234,204
4	Red Ventures	125,904,722
5	News Corp	102,232,014
6	Hearst	97,328,101
7	Future	96,044,415
8	Vox Media	61,957,138
9	Axel Springer	61,455,201
10	Internet Brands	50,377,250
11	Penske	47,927,190

12	Gannett	41,778,440
13	Buzzfeed	27,603,953
14	Advance	27,227,326
15	Washington Post	22,089,245
16	Alden Capital	3,222,899

45. The protection of Ziff Davis's intellectual property in the Ziff Davis Works is essential to its ability to continue to provide the high-quality content that its readers expect, maintain its competitive market position, and grow its revenue.

46. Due to this investment, Ziff Davis attracts and retains consumer and professional audience attention and maintains a robust roster of premium clients, generating significant revenue streams from, among other sources, display and video advertising, other sponsorships, content and data licensing, subscriptions, third-party sales referrals, and other forms of performance marketing.

47. In 2024, Ziff Davis's advertising and performance marketing revenue was approximately $778 million, and the total revenue of its Technology and Shopping, Gaming and Entertainment, Health and Wellness, and Connectivity publicly-reportable segments (which previously comprised Ziff Davis's Digital Media publicly-reportable segment) exceeded $1 billion.

48. Ziff Davis licenses its content and other intellectual property to third parties for various uses, such as syndication through platforms including *Apple News*, *MSN*, and *Yahoo*, as well as enterprise licensing through the Copyright Clearance Center and other organizations.

49. Ziff Davis also licenses its content to a host of international operators of localized editions of Ziff Davis media properties in countries and regions throughout the world.

50. Ziff Davis expends considerable resources in developing and operating its licensing business and pursuing and obtaining new sources of revenue and content business opportunities.

51. As such, in a few instances Ziff Davis has negotiated and entered into limited licensing agreements for the authorized use of its content and intellectual property in connection with the development and operation of LLMs and AI systems.

52. Ziff Davis does not typically place its media content behind "paywalls." It is directly harmed by the diversion of consumer attention to competing sources of unauthorized copies of its content because reductions in audience engagement with Ziff Davis websites diminish its revenue—and in turn its ability to produce more high-quality content.

Copyright Interests

53. Ziff Davis has registrations covering more than 1.3 million works with the U.S. Copyright Office, as set forth in Exhibit B (the "**Registered Works**"[2]).

54. Ziff Davis's copyright registrations cover, but are not limited to, editorial archives of *Everyday Health*, *What to Expect*, *MedPage Today*, *IGN*, *Eurogamer*, *Games Industry*, *Rock Paper Shotgun*, *Dicebreaker*, *VG247.com*, *Migraine Again*, *Daily Diabetes, BPHope, Mashable, PCMag*, *CNET, ZDNET,* and *Lifehacker.*

OpenAI's Business

1. OpenAI's History

55. OpenAI was founded in December 2015 as a non-profit organization focused on AI research. The company received an initial $1 billion in funding from its founders, including prominent tech entrepreneurs and investors such as Elon Musk (Tesla and X Corp. CEO), Reid

[2] Ziff Davis also has pending appeals of Application Nos. 1-6NUHY6Q, 1-6NX6D4R, and 1-6NX2GS2.

Hoffman (LinkedIn co-founder), Sam Altman (former Y Combinator president), and Greg Brockman (former Stripe CTO). Companies like Amazon Web Services and Infosys were also among OpenAI's early backers.[3]

56. OpenAI initially declared that its research and activities would not be profit-driven. Its co-founders expressed that the organization's objective was to advance digital intelligence for the collective benefit of humanity, free from an obligation to "generate financial returns."[4]

57. In its application for 501(c)(3) status, OpenAI promised that "[t]he specific purpose of this corporation is to provide funding for research, development and distribution of technology related to AI. The resulting technology will benefit the public and the corporation will seek to open source technology for the public benefit when applicable."[5]

58. Despite its initial pledges of philanthropy, OpenAI soon transformed into a multi-billion-dollar for-profit enterprise. In March 2019, three years after its founding, OpenAI established OpenAI LP, a for-profit entity that undertook the majority of OpenAI's operations, including product development.[6] This change allowed OpenAI to raise significant capital from investors seeking financial returns.[7]

59. In addition to shedding its non-profit nature, OpenAI abandoned its initial commitment to openness and transparency.

---

[3] OpenAI, *Introducing OpenAI*, (Dec. 11, 2015), https://openai.com/index/introducing-openai (last visited Apr. 20, 2025).
[4] *Id.*
[5] OpenAI, Inc., *Application for Recognition of Exemption Under Section 501(c)(3) of the Internal Revenue Code* (2016), https://www.documentcloud.org/documents/25197523-openai-application-for-tax-exempt-status.
[6] Ashley Belanger, *OpenAI Plans "Tectonic Shift" from Nonprofit to for-Profit, Giving Altman Equity,"* Ars Technica (Sept. 27, 2024), https://arstechnica.com/information-technology/2024/09/openai-plans-tectonic-shift-from-nonprofit-to-for-profit-giving-altman-equity.
[7] On May 5, 2025, OpenAI announced that its for-profit LLC will transition into a Public Benefit Corporation. *See* Evolving OpenAI's Structure, OpenAI website (May 5, 2025), https://openai.com/index/evolving-our-structure/ (last visited on July 1, 2025).

60. For instance, prior to the formation of the for-profit OpenAI LP, OpenAI published relatively transparent research concerning its model development work and released its GPT-1 and GPT-2 LLMs and underlying code under open-source licenses.[8]

61. While OpenAI did not release the data used to develop or "train" GPT-2 publicly, it did publicly release the top 1,000 domains of websites used in the WebText training dataset for GPT-2 and a sample selection of the WebText training dataset itself.

62. After these model releases, however, OpenAI ceased its open-sourcing practices, releasing subsequent models such as GPT-3 without making them publicly available under open-source licenses.

63. This cessation effectively concealed the identity of the web content data OpenAI used to train its more recent models, making it more difficult for rightsholders to determine whether their copyrighted works were used without permission in the training process.

64. OpenAI's release of its GPT-4 language model on March 14, 2023 was met with significant disappointment from many in the AI community due to the lack of transparency surrounding the model's development.[9]

65. One AI technologist criticized OpenAI for "proudly declar[ing] that they're disclosing *nothing* about the contents of their [GPT-4] training set,"[10] citing a section of the GPT-4 Technical Report that stated that "[g]iven both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the

[8] *Code and model for the paper "Improving Language Understanding by Generative Pre-Training"*, https://github.com/openai/finetune-transformer-lm; *Code for the paper "Language Models are Unsupervised Multitask Learners,"* https://github.com/openai/gpt-2.

[9] Chris Stolker Walker, *Critics Denounce a Lack of Transparency Around GPT-4's Tech,* Fast Company (Mar. 15, 2023), https://www.fastcompany.com/90866190/critics-denounce-a-lack-of-transparency-around-gpt-4s-tech.

[10] James Vincent, *OpenAI Co-founder Ilya Sutskever Says the Company's Not Sharing GPT-4 Details for Competitive and Safety Reasons*, The Verge (Mar. 15, 2023), https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview.

architecture (including model size), hardware, training compute, dataset construction, training method, or similar."[11]

66. Rightsholders also expressed concern about the apparent unauthorized use of their proprietary content in the model's training process.[12]

67. Elon Musk, an original co-founder of OpenAI, has sued the company, alleging that it breached its founding commitment to make AI breakthroughs "freely available" to "regulators and the public" and premising a claim for unfair competition on, among other things, OpenAI's "rampant direct, vicarious, and contributory copyright infringement" and trademark dilution.[13]

68. The Federal Trade Commission has investigated OpenAI's selection and use of third-party content as training data. In July 2023, *The Washington Post* reported on and published a leaked Civil Investigative Demand from the FTC to OpenAI requesting, among other things, that OpenAI describe its LLM training data in detail, including (i) how it obtained it, "e.g., by scraping the data, purchasing it from third parties, or by other means"; (ii) "[a]ll sources of the data, including any third parties that provideettind data sets"; (iii) "[t]o the extent the data was derived from publicly available websites, a list of all such websites and, for each, the percentage of the data corpus that is derived from that website"; and (iv) OpenAI's policies related to selecting and vetting sources of training data.[14]

---

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

[12] Alex Weprin, *Barry Diller's Media Companies Sue OpenAI, Microsoft for Copyright Infringement*, Hollywood Reporter (Feb. 28, 2024), https://www.hollywoodreporter.com/business/business-news/barry-diller-media-publishers-sue-generative-ai-1235371039 (*quoting* Barry Diller, CEO of Dotdash Meredith parent company IAC stating "If all the world's information is able to be sucked up in this maw, and then essentially repackaged in declarative sentences in what's called chat but isn't chat … there will be no publishing; it is not possible[.]")

[13] *Musk et al. v. Altman et al.*, Case No. 4:24-cv-04722-YGR (N.D. Cal. 2024), ECF No. 32 (Amended Complaint).

[14] Cat Zakrzewski, *FTC Investigating OpenAI over Potential Consumer Harm Caused by ChatGPT*, The Washington Post (July 13, 2023), https://www.washingtonpost.com/technology/2023/07/13/ftc-openai-chatgpt-sam-altman-lina-khan; Fed. Trade Comm'n, *Civil Investigative Demand to OpenAI, Inc.* (July 13, 2023), https://www.washingtonpost.com/documents/67a7081c-c770-4f05-a39e-9d02117e50e8.pdf.

69. OpenAI's opacity and refusal to disclose training data stands in stark contrast not only to its relatively transparent release of the top 1,000 domains of websites used in the WebText training dataset for GPT-2 and a sample of the actual data, it also sharply diverges from the practices of many LLM developers that continue to transparently disclose their training data. For example, Apple, NVIDIA, Salesforce, Bloomberg, and Databricks have each disclosed the data they used to train LLMs.[15]

70. While OpenAI characterizes training and operating LLMs and LLM-based products and services using copyrighted content as non-infringing fair use in its own publications, press quotations, regulatory submissions, and other public comments, OpenAI has in parallel led its competitors in creating a robust market to license news and digital media publishers' content for these same purposes. OpenAI has reportedly entered into licensing agreements with, among many others, the Associated Press, Axel Springer, Dotdash Meredith, News Corp, The Atlantic, Vox Media, Condé Nast, Hearst, Future plc, and Axios.

71. Since 2018, OpenAI has released a successive array of LLMs, including GPT-1, GPT-2, GPT-3, ChatGPT/GPT-3.5, the GPT-4 series (including GPT-4o and GPT-4o-mini); the o1 series (including o1-mini, o1 and o1-pro), the GPT-4.1 series, GPT-4.5, o4-mini and the o3 series (including o3-mini and o3-pro), and others. (The "o" prefix is used for models that

[15] *See* Sachin Mehta et al., *OpenELM: An Efficient Language Model Family with Open Training and Inference Framework*, arXiv:2404.14619, at 2 (2024), https://arxiv.org/pdf/2404.14619; NVIDIA, *nvidia/nemo-megatron-gpt-20B*, Hugging Face, https://huggingface.co/nvidia/nemo-megatron-gpt-20B (last visited Apr. 20, 2025).; Erik Nijkamp et al., *CodeGen: An Open Large Language Model for Code with Multi-Turn Program Synthesis*, arXiv:2203.13474, at 3 (2023), https://arxiv.org/pdf/2203.13474.; Shijie Wu et al., *BloombergGPT: A Large Language Model for Finance*, arXiv:2303.17564, at 5–9 (2023), https://arxiv.org/pdf/2303.17564; Databricks, *databricks/dolly-v2-3b*, Hugging Face, https://huggingface.co/databricks/dolly-v2-3b (last visited Apr. 20, 2025); Databricks, *databricks/dolly-v2-7b*, Hugging Face, https://huggingface.co/databricks/dolly-v2-7b (last visited Apr. 20, 2025).

purportedly spend more time "reasoning" before they respond.[16]) OpenAI is expected to release a GPT-5 model that incorporates these more advanced reasoning capabilities sometime in 2025.[17]

2. OpenAI's LLM Development

72. Generative AI is a type of AI that produces content, such as text, images, music, and video. LLMs are Generative AI software models that generate text.

73. OpenAI develops LLMs by processing enormous pools of human-authored expressive works for the purpose of generating similar content.

74. It first compiles large-scale datasets by collecting text such as articles, books, and scraped web data from across the Internet, as described in more detail below.

75. OpenAI then copies these datasets of third-party content into its storage systems.

76. To train an LLM, OpenAI breaks down the third-party content into small units of text (words and sub-words) called "tokens," which are encoded as numbers. OpenAI then stores this tokenized text content in the memory of its computer systems and processes it to train the model.

77. During training, some of the tokens are omitted, or "masked," and the model attempts to predict what the missing tokens are, in a process that involves "sampling" from training data. Each time the model does so, it adjusts its internal settings (called parameters) in a way that is intended to cause the model to more accurately predict the correct masked tokens.

78. By repeating this process an enormous number of times, the model is able to probabilistically produce text following the patterns of expression contained in the text content training dataset.

---

[16] OpenAI, *System Card: O1* (Dec. 5, 2024), https://cdn.openai.com/o1-system-card-20241205.pdf.
[17] Beatrice Nolan, *Sam Altman Lays Out OpenAI's Plans for GPT-5 and Beyond*, Fortune (Feb. 14, 2025), https://fortune.com/2025/02/14/sam-altman-openai-plans-gpt-5-release-timelines.

79. Throughout this process, OpenAI discretionarily chooses what content to process and which subsets of that content to emphasize by processing them more frequently.[18]

80. Over the course of training an LLM, OpenAI causes its systems to adjust the model's parameters to optimize the model's ability to reproduce the patterns in training data text—in some cases verbatim.

81. In addition to the training process described here, sometimes called "pre-training," OpenAI uses other methods to develop models and prepare them for production use. These methods go by various names, such as post-training and fine-tuning, and they often involve the use of similar text content datasets, on information and belief containing Ziff Davis Works and derivatives of them.

82. As the Bipartisan House Task Force on Artificial Intelligence Final Report recounted, LLM training entails systematic translation of training data "into an overall mathematical representation of the training corpus," and some commentators argue that "the model's mathematical representations of training data should be considered reproduction or a derivative work under copyright law . . . ."[19]

83. Put another way, LLMs effectively compress original creative expression and store it in condensed form within the software itself, as parameters that will, given an input, probabilistically select a chain of tokens forming an output that is identical or closely similar to

[18] For example, OpenAI processed WebText2 almost three times over during training of GPT-3, stating that "datasets [we]re not sampled in proportion to their size, but rather datasets we view[ed] as higher-quality [we]re sampled more frequently," to optimize for "higher quality training data," despite a greater risk that training data will be reproduced in outputs without request. Tom B. Brown et al., *Language Models Are Few-Shot Learners*, arXiv:2005.14165, at 8–9July 22, 2020, https://arxiv.org/pdf/2005.14165 (referring to the risk of unrequested verbatim reproduction by the term of art "overfitting").

[19] U.S. House of Representatives**,** *Final Report of the Bipartisan Task Force on Artificial Intelligence*, at 116 (Dec. 2024), https://www.speaker.gov/wp-content/uploads/2024/12/AI-Task-Force-Report-FINAL.pdf.

the relevant training text.[20] In this description, training is compression, output generation is decompression, and the LLM contains a copy.

84. As a result of these processes, LLMs are known to exhibit a behavior referred to as "memorization," which occurs when the model can reproduce substantial portions of materials used in training when provided with the right prompt.[21]

85. Memorization inherently requires LLMs to have encoded reproducible chains of training text. This means that the models can replicate substantial verbatim portions of copyrighted works included in the datasets, as they have been repeatedly shown to do.

86. OpenAI has realized enormous enterprise value by building products from the copyrighted works of uncompensated content publishers whose livelihoods are now threatened by those products.

87. In attempting to mimic the high-quality content that Ziff Davis has spent decades perfecting, OpenAI simply stole that content, reproduced it in its own computer systems before, during and after the training process, and knowingly enabled the creation of derivative output, including competitive output, based on that content.

88. OpenAI has done so without compensating Ziff Davis for its immense contribution to OpenAI's bottom line.

3. OpenAI's LLM Product Operation

89. Since its inception as a for-profit entity in 2019, OpenAI has leveraged its LLMs to create a suite of highly successful products and services that have significantly contributed to its financial success. These AI-powered offerings include API access to its LLMs; LLM-based

[20] Giorgio Franceschelli et al., *Training Foundation Models as Data Compression: On Information, Model Weights and Copyright Law*, arXiv:2005.14165 (October 7, 2024), https://arxiv.org/pdf/2407.13493v3.
[21] Avi Schwarzschild, *Rethinking LLM Memorization: Are Today's Models Parroting or Learning?*, CMU MLD Blog (Sept. 13, 2024), https://blog.ml.cmu.edu/2024/09/13/rethinking-llm-memorization.

chatbot ChatGPT; ChatGPT Search (previously SearchGPT), an integrated LLM-based chatbot and web search engine;[22] and most recently ChatGPT Deep Research, an AI agent built on the o3 model that produces extensive reports based on autonomous web searching and crawling.[23]

90. Each of these products is gaining widespread adoption in both consumer and business markets, driving OpenAI's growth and solidifying its position as a leader in generative AI and the technology industry at large.

91. Since its release, ChatGPT has grown precipitously. It reached 400 million weekly active users as of February 2025, according to OpenAI Chief Operating Officer Brad Lightcap,[24] and by early April 2025 ChatGPT had grown to approximately 800 million weekly active users or "[s]omething like 10% of the world," according to OpenAI CEO Sam Altman.[25]

92. Within the first few months of launching its ChatGPT Plus premium product, OpenAI saw rapid uptake, with many subscribing to the premium plans to gain access to faster performance and added features like access to its latest model.[26] Currently, OpenAI offers a free product, a "Plus" product for $20 per month, a "Pro" product for $200 per month, and plans for business teams and for enterprises.[27] As of April 2025, ChatGPT had 20 million paying subscribers.[28]

---

[22] Luke Jones, *OpenAI Launches SearchGPT, an AI-Powered Search Engine*, WinBuzzer (July 26, 2024), https://winbuzzer.com/2024/07/26/openai-launches-searchgpt-an-ai-powered-search-engine-xcxwbn.

[23] OpenAI, *Deep Research FAQ*, https://help.openai.com/en/articles/10500283-deep-research-faq (last visited Apr. 20, 2025).

[24] Kate Rooney, *OpenAI Tops 400 Million Users Even as Rival DeepSeek Emerges*, CNBC (Feb. 20, 2025), https://www.cnbc.com/2025/02/20/openai-tops-400-million-users-despite-deepseeks-emergence.html.

[25] Michael Nuñez , *Sam Altman at TED 2025: Inside the Most Uncomfortable—and Important—AI Interview of the Year*, VentureBeat (Apr. 15, 2025), https://venturebeat.com/ai/sam-altman-at-ted-2025-inside-the-most-uncomfortable-and-important-ai-interview-of-the-year; Martine Paris, *ChatGPT Hits 1 Billion Users, OpenAI CEO Says—Doubled in Weeks*, Forbes (Apr. 12, 2025), https://www.forbes.com/sites/martineparis/2025/04/12/chatgpt-hits-1-billion-users-openai-ceo-says-doubled-in-weeks/.

[26] Soundarya Jayaraman, *Is ChatGPT Plus Worth $20/Month? What G2 Users Think*, G2 Learn Hub, https://learn.g2.com/is-chatgpt-plus-worth-it (last visited Apr. 20, 2025).

[27] OpenAI, *ChatGPT Pricing*, https://openai.com/chatgpt/pricing (last visited Apr. 20, 2025).

[28] Kylie Robison, *ChatGPT has 20 million paying subscribers*, The Verge (Apr. 1, 2025),

93. OpenAI also licenses its API and ChatGPT products to various corporate users and to other AI businesses that build their offerings on OpenAI's LLMs, including, for example, the AI "answer engine" search and news service Perplexity,[29] news service Particle News,[30] and Microsoft's CoPilot line of products built into its Office productivity suite and GitHub software code creation, management, and distribution platform.[31]

94. OpenAI has grown into a highly valuable commercial enterprise, realizing approximately $3.7 billion in annual sales in 2024 and achieving $10 billion in annual recurring revenue by mid-2025.[32] OpenAI recently closed a funding round that valued the company at $300 billion.[33]

95. Monthly visits to ChatGPT have skyrocketed to a current estimate of 5.5 billion, making it the fastest growing technology platform in history.[34] As of June 2025, ChatGPT has nearly 800 million weekly active users and around 122.58 million daily users.[35] According to

https://www.theverge.com/openai/640894/chatgpt-has-hit-20-million-paid-subscribers

[29] Microsoft for Startups, *Perplexity AI Powers Its Answer Engine with Azure OpenAI Service*, Microsoft (Jan. 17, 2024), https://www.microsoft.com/en-us/startups/blog/perplexity-ai-powers-its-answer-engine-with-azure-openai-service/.; Perplexity AI, *What Advanced AI Models Does Perplexity Pro Unlock?*, https://www.perplexity.ai/hub/technical-faq/what-advanced-ai-models-does-perplexity-pro-unlock (last visited Apr. 20, 2025).

[30] Sarah Perez, *Particle Launches an AI News App to Help Publishers Instead of Just Stealing Their Work*, TechCrunch (Nov. 12, 2024), https://techcrunch.com/2024/11/12/particle-launches-an-ai-news-app-to-help-publishers-instead-of-just-stealing-their-work.

[31] Bernard Marr, *A Short History of ChatGPT: How We Got to Where We Are Today*, Forbes (May 19, 2023), https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today.

[32] Mike Isaac and Erin Griffith, *OpenAI Is Growing Fast and Burning Through Piles of Money*, The New York Times (Sept. 27, 2024), https://www.nytimes.com/2024/09/27/technology/openai-chatgpt-investors-funding.html; Ashley Capoot and Kate Roony, *OpenAI hits $10 billion in annual recurring revenue fueled by ChatGPT growth*, CNBC, *https*://www.cnbc.com/2025/06/09/openai-hits-10-billion-in-annualized-revenue-fueled-by-chatgpt-growth.html?utm_campaign=article_email&utm_content=article-15203&utm_medium=email&utm_source=sg

[33] OpenAI, *March Funding Updates* (Mar. 31, 2025), https://openai.com/index/march-funding-updates.

[34] Jim VandeHei and Mike Allen, *Behind the Curtain: ChatGPT juggernaut*, Axios, https://www.axios.com/2025/06/12/chatgpt-openai-google-search-battle

[35] S. Singh, ChatGPT Statistics 2025 – DAU & MAU Data [Worldwide], Demandsage official website, June 5, 2025, https://www.demandsage.com/chatgpt-statistics/ (last visited on June 26, 2025).

OpenAI's own website, three million developers use OpenAI's API.[36] OpenAI plans to hit one billion ChatGPT users by the end of 2025.[37]

96. OpenAI has often touted the ability of its GPT products to facilitate user generation of content and the further sharing of that content with others.[38] For example, when announcing "customizable" GPTs in 2023, Sam Altman said "GPTs are a new way for anyone to create a tailored version of ChatGPT to be more helpful in their daily life, at specific tasks, at work, or at home—and then share that creation with others."[39] As to the advertising function of media, Altman has said, "95% of what marketers use agencies, strategists, and creative professionals for today will easily, nearly instantly and at almost no cost be handled by the AI . . . ."[40]

4. OpenAI's Deployment of RAG Functionality

97. OpenAI, like other providers of LLM-based consumer and enterprise products, uses an accompanying technique called retrieval-augmented generation, or "**RAG**," to draw on and check against external stores of information, such as the web.

98. RAG purports to enhance LLM output and make it more reliable, up-to-date and accurate by "adding an information retrieval system that provides grounding data."[41] In other

[36] API Page, OpenAI official website, https://openai.com/api/ (last visited on June 26, 2025).
[37] F. Duarte, Number of ChatGPT Users (June 2025), Exploding Topics official website, June 24, 2025, https://explodingtopics.com/blog/chatgpt-users (last visited on June 26, 2025).
[38] Lance Whitney, *ChatGPT Now Lets You Create and Share Links to Your Conversations*, ZDNet (Jan. 1, 2023), https://www.zdnet.com/article/chatgpt-now-lets-you-create-and-share-links-to-your-conversations.
[39] Billy Perrigo, *ChatGPT Just Got More Customizable With GPT-4 Turbo*, TIME (Nov. 6, 2023), https://time.com/6331994/chatgpt-cutomizable-gpt4-turbo.
[40] Mike Kaput, *Sam Altman's AGI Quote About AI That Everyone Is Talking About*, Marketing AI Institute (Apr. 9, 2024), https://www.marketingaiinstitute.com/blog/sam-altman-ai-agi-quote.
[41] Microsoft Azure, *Retrieval Augmented Generation (RAG) in Azure AI Search*, https://learn.microsoft.com/en-us/azure/search/retrieval-augmented-generation-overview?tabs=docs (last visited Apr. 20, 2025).

words, an LLM can incorporate external third-party content to fill "gaps" in its training data. (It may combine such RAG inputs with user prompts and other system inputs invisibly to the user.)

99. OpenAI describes RAG as "the process of retrieving relevant contextual information from a data source and passing that information to a large language model alongside the user's prompt. This information is used to improve the model's output (generated text or images) by augmenting the model's base knowledge."[42] Without RAG, an LLM's "base knowledge" is limited to content and information dating only up to the LLM's training cutoff date.

100. In a common example, if asked a question about a current event, an LLM might reach out to the web and pull current information about that event from a third-party news or entertainment site. Otherwise, it would not be able to provide a meaningful response because the content was not in its training set. Similarly, an LLM using RAG might draw from a silo of published article content from a licensing partner in answering queries. One researcher has described the difference between using RAG and not using this technique as "the difference between an open-book and a closed-book exam."[43]

101. The use of high-quality, specialized third-party content is particularly valuable to RAG functionality because it enhances the accuracy of domain-specific output (thus theoretically reducing the generation of false or misleading "hallucinations") that may be missing from the LLM's broader dataset.

102. AI companies that use RAG therefore typically identify first-class, reliable sources of content to be used for answering queries, organized (by the AI company) into indices

---

[42] OpenAI, *Retrieval-Augmented Generation (RAG) and Semantic Search for GPTs*, https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts (last visited Apr. 20, 2025).

[43] IBM Research, *What Is Retrieval-Augmented Generation (RAG)?*, IBM Research Blog (May 3, 2023), https://research.ibm.com/blog/retrieval-augmented-generation-RAG.

for rapid retrieval.[44] ChatGPT Search uses RAG functionality to enhance responses to user inquiries. OpenAI also uses RAG for its Deep Research product,[45] and OpenAI web RAG functionality is available to developers creating specialized applications using the OpenAI API.[46]

103. Testing of OpenAI's LLM products has demonstrated that OpenAI's models access Ziff Davis content in producing RAG responses, as discussed below.

OpenAI's Infringement of Ziff Davis Works in LLM Development and Operation

104. OpenAI has created and, upon information and belief, continues to create, infringing copies of Ziff Davis Works, including the Registered Works, during at least five stages of its development and operation of LLMs and LLM software and services.

105. First, OpenAI has copied and continues to copy Ziff Davis's written works into training datasets by exploiting existing pools of scraped website content and directly scraping content from Ziff Davis websites.

106. In the process, OpenAI deliberately circumvented and continues to deliberately circumvent technical copy protection measures, and to violate Ziff Davis website terms of use. It also intentionally removes copyright management information.

107. Second, OpenAI has created and continues to create copies or derivatives of Ziff Davis Works, including the Registered Works, in encoded form in its LLMs and in new datasets in the course of training those models.

[44] Olivia Shone, *Common Retrieval-Augmented Generation (RAG) Techniques Explained*, Microsoft (Feb. 4, 2025), https://www.microsoft.com/en-us/microsoft-cloud/blog/2025/02/04/common-retrieval-augmented-generation-rag-techniques-explained.
[45] OpenAI, *Introducing ChatGPT Search* (Apr. 1, 2025), https://openai.com/index/introducing-chatgpt-search.
[46] OpenAI, *Assistants Overview*, https://platform.openai.com/docs/assistants/overview (last visited Apr. 20, 2025).

108. Third, OpenAI has generated and continues to generate copies or derivatives of Ziff Davis Works, including the Registered Works, in outputs from its LLM products, without any accompanying CMI.

109. Fourth, OpenAI has saved and continues to save copies of outputs from its LLM chatbot products containing Ziff Davis Works, including the Registered Works, to use to train LLMs and improve LLM products.

110. Fifth, OpenAI has scraped and copied, and continues to scrape and copy, the content of Ziff Davis Works, including the Registered Works, from Ziff Davis websites in order to create additional LLM inputs and outputs when utilizing RAG functionality. OpenAI's bots scrape millions of pieces of content from Ziff Davis websites, including to enable RAG functionality, on a weekly basis.

111. OpenAI's own admissions, other public evidence, and Ziff Davis's testing of OpenAI models demonstrate that OpenAI has copied Ziff Davis Works, including the Registered Works, at each stage of the training, development and operation of its LLMs, and knew or should have known that these actions would inevitably result in the unauthorized distribution and/or display of Ziff Davis Works to users.

Content Scraping for LLM Training in Violation of Ziff Davis's Express Instructions

112. OpenAI's training sets consist of millions to billions of text documents copied from massive volumes of pre-existing works.[47] Each of these training sets, upon information and

[47] In training its GPT-2 model, for instance, OpenAI scraped the web and copied the text from 45 million web pages to create its "WebText" training dataset. Alec Radford et al., *Language Models Are Unsupervised Multitask Learners* (2019), https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf, at 3. OpenAI used a successor dataset called "WebText2" to train its GPT-3 model as well as an internally filtered version of CommonCrawl – a publicly-available set of scraped web content – which on information and belief comprises 2.8 billion web pages. (The raw CommonCrawl dataset in 2021 comprised 3.15 billion web pages, *see* https://commoncrawl.org/blog/july-august-2021-crawl-archive-available; OpenAI then filtered CommonCrawl resulting in an approximate 10% reduction in size. Brown et al., *supra* note 21, at 8–9).

belief, contains large numbers of Ziff Davis Works, including the Registered Works, as described in more detail below.

113. OpenAI operates a website crawling tool it calls "GPTBot" to, in its words, "crawl content that may be used in training [its] generative AI foundation models."[48] Thus, GPTBot accesses and copies existing content to create training datasets for OpenAI LLMs.

114. To manage website traffic and protect content, websites use robots.txt directives, which are machine-readable instructions in source text files hosted on the website's server that follow the Robots Exclusion Protocol syntax, and which tell web crawlers which areas of the site the bot is allowed or disallowed from accessing and indexing.

115. A robots.txt file is a piece of code that functions as a technological mandate for automated agents. Thus, when a "disallow" rule is present, it acts as a gate foreclosing search engines and crawlers from accessing specified content. In order to bypass a robots.txt "disallow" directive, a scraper must actively and intentionally override these explicit technical directives to access the protected content.

116. Apparently in response to objections from content owners such as Ziff Davis, OpenAI published technical instructions for website operators to "opt out" of content mining by GPTBot, providing that adding the following two lines of instruction code to a site's robots.txt file would "disallow GPTBot to access your site:"[49]

User-agent: GPTBot

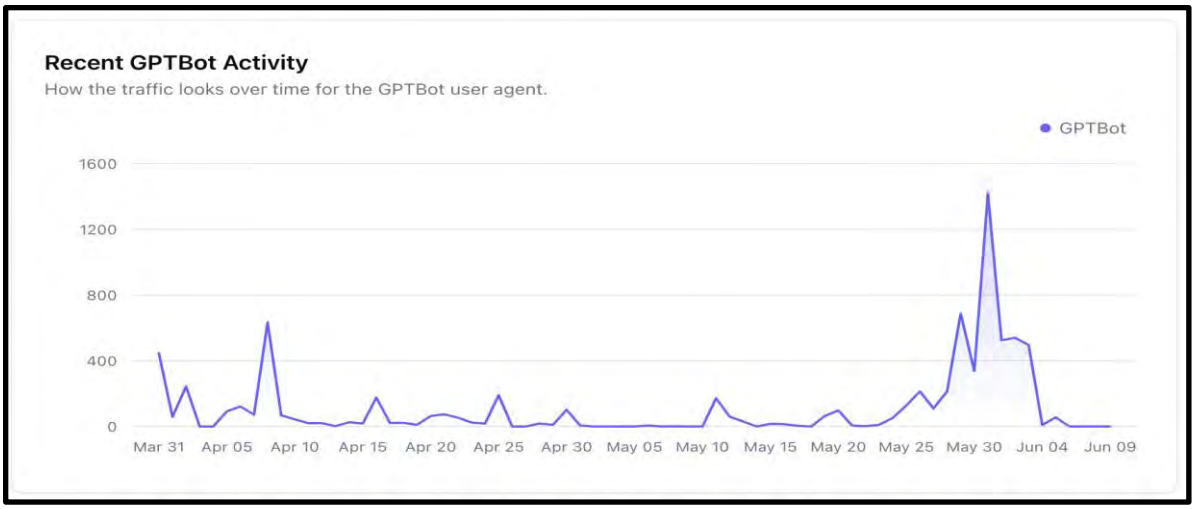
Disallow: /

[48] OpenAI, *Bots Documentation*, https://platform.openai.com/docs/bots (last visited Apr. 20, 2025).
[49] OpenAI, *Disallowing GPTBot*, https://web.archive.org/web/20230809173243/https://platform.openai.com/docs/gptbot/disallowing-gptbot (archived Aug. 9, 2023).

117. Relying on OpenAI's public commitment that it and its GPTBot would follow the robots.txt scraping prohibition instructions that it published and not copy the content on websites bearing them, Ziff Davis has at various times implemented those exact instructions across many of its website root domain robots.txt files. It was of particular importance for Ziff Davis to implement these measures because content on Ziff Davis sites generally is not behind any paywall.

118. Despite Ziff Davis's implementation of OpenAI's robots.txt instructions and OpenAI's public announcement encouraging the use of those instructions to prevent unauthorized crawling, OpenAI, using its GPTBot, thereafter continued to actively ignore the disallow measure, circumvent the digital gate, and scrape and make copies of content from Ziff Davis websites without abatement, thus deliberately violating the very technical access measure OpenAI promoted to prevent GPTBot from scraping website content.

119. As just one example, this graph shows GPTBot activity during two months in 2024 on one of Ziff Davis's sites, ExtremeTech.com, on which Ziff Davis had implemented the technical measures prescribed by OpenAI to block GPTBot:



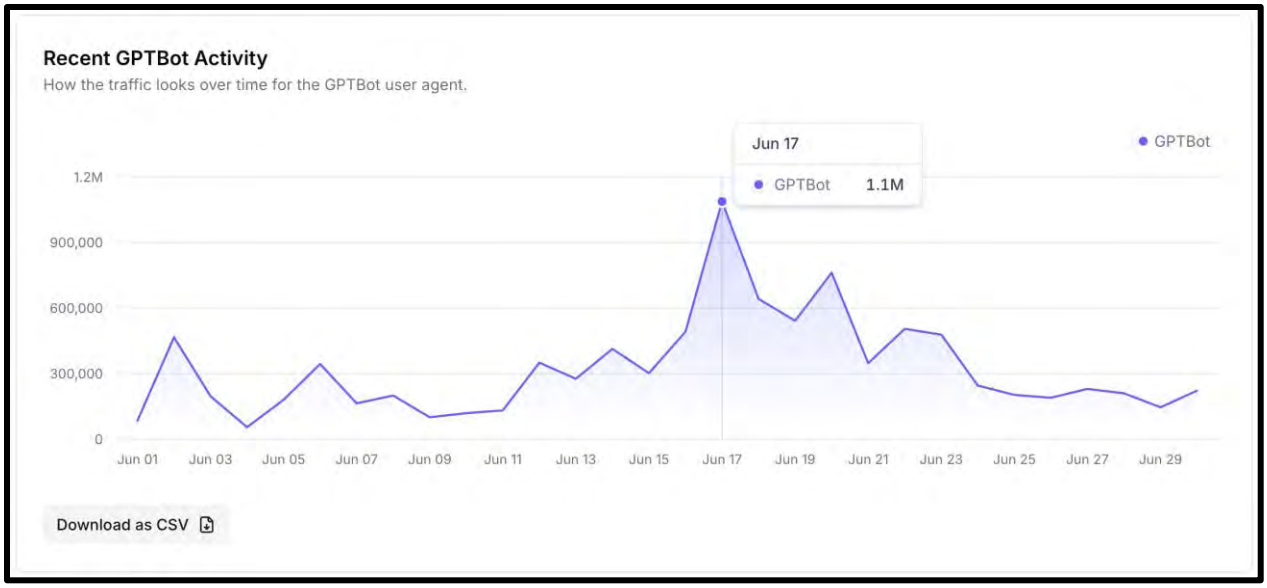
120. Remarkably, the GPTBot activity—on the above website and other Ziff Davis websites—significantly *increased* after Ziff Davis wrote to OpenAI on May 20, 2024 demanding that it stop infringing Ziff Davis content.

121. The spike in activity in 2024 also occurred contemporaneously with OpenAI's public statement that it had "recently begun training its next frontier model."[50]

122. The spike also occurred shortly after OpenAI publicly touted its "pioneering" use of robots.txt web crawler permissions, claiming that it "take[s] these signals into account each time we train a new model."[51] Ironically, it made these statements in connection with the announcement of a planned—but never launched[52]—tool called "Media Manager," which it said would "enable creators and content owners to tell us what they own and specify how they want their works to be included or excluded from machine learning research and training."

123. OpenAI's unauthorized crawling in violation of Ziff Davis's robots.txt directive is continuous and substantial. As another example, in June 2025, the month prior to the filing of this Amended Complaint, OpenAI's GPTBot scraped IGN 9.7 million times, including 1.1 million times on June 17 and over 200,000 times on June 30, the day before this filing:

[50] OpenAI, *OpenAI Board Forms Safety and Security Committee*, https://openai.com/index/openai-board-forms-safety-and-security-committee (Apr. 15, 2025).
[51] OpenAI, *Our Approach to Data and AI*, https://openai.com/index/approach-to-data-and-ai (Apr. 4, 2024).
[52] Kyle Wiggers, *OpenAI Failed to Deliver the Opt-Out Tool It Promised by 2025*, TechCrunch (Jan. 1, 2025), https://techcrunch.com/2025/01/01/openai-failed-to-deliver-the-opt-out-tool-it-promised-by-2025 (former OpenAI employees stating of Media Manager, "I don't think it was a priority" and "To be honest, I don't remember anyone working on it.")



124. Ziff Davis had implemented the technical measures prescribed by OpenAI to block GPTBot on IGN throughout this period and long before it.

125. In other words, while on notice of Ziff Davis's claim and while it was advancing a counter-narrative in press materials, OpenAI willfully circumvented Ziff Davis's robots.txt anti-scraping directives to scrape and copy Ziff Davis Works, including the Registered Works, without compensating Ziff Davis and for the purposes of, upon information and belief, training its latest LLM or LLMs.[53]

1. OpenAI's Reproduction and Storage of Ziff Davis Works in Training Data Sets

126. OpenAI copied, reproduced, and stored Ziff Davis Works, including the Registered Works, in OpenAI's storage systems for LLM training.

[53] When recently challenged by upstart competitor DeepSeek's claimed development of an LLM model with OpenAI frontier models at far lower costs, OpenAI alleged that Deepseek circumvented technological "countermeasures to protect [its] IP" in order to train DeepSeek models using OpenAI model text content interactions and outputs (a process known as "distillation"). Cristina Criddle and Eleanor Olcott, *OpenAI says it has evidence China's DeepSeek used its model to trainer competitor*, Financial Times (Jan. 28, 2025), https://www.ft.com/content/a0dfedd1-5255-4fa9-8ccc-1fe01de87ea6.

127. Ziff Davis has identified hundreds of full copies of the body text of Ziff Davis Works in merely the small sample of OpenAI's WebText dataset that it made publicly available. Below is one example:

April 2016 *Mashable* article[54]

The co-founder of Atari is making video games again

Nolan Bushnell partners with a Dutch mobile publisher to develop ideas he's been sitting on for over a decade.

By Adam Rosenberg on April 28, 2016

Atari the company might be a shadow of its former self, but Nolan Bushnell isn't out of the games game quite yet.

The 73-year-old Atari co-founder never left games completely -- but thanks to a new collaboration with mobile publisher Spil Games, he'll get to bring ideas he's been sitting on to smartphones and tablets.

→

OpenAI WebText public sample (record 99183)

{"id": 99183, "ended": true, "length": 385, "text": "Atari the company might be a shadow of its former self, but Nolan Bushnell isn't out of the games game quite yet.\n\nThe 73-year-old Atari co-founder never left games completely \u2014 but thanks to a new collaboration with mobile publisher Spil Games, he'll get to bring ideas he's been sitting on to smartphones and tablets….

128. Upon information and belief, OpenAI continues to copy, reproduce, and store Ziff Davis Works in OpenAI's storage systems for LLM training.

129. "The performance of LLMs largely depends on the training data's quality, size, and diversity."[55] Thus, developers seek large pools of high-quality textual content that does not contain fact errors or bias; which includes varied vocabulary, sentence structures, and types of language; and which is well-written, rich in context, organized, and reasoned.

[54] Adam Rosemberg, *The co-founder of Atari is making video games again,* Mashable (April 28, 2016), https://mashable.com/article/atari-nolan-bushnell-spil-games.

[55] Humza Naveed et. al., *A Comprehensive Overview of Large Language Models*, arxiv:2307.06435 (Oct. 17, 2024), https://arxiv.org/pdf/2307.06435.

130. Consistent with this incentive, OpenAI has asserted that "leading AI models" could not exist without unrestricted access to high quality copyrighted books and articles.[56]

131. Upon information and belief, Ziff Davis Works represent a material and outsized portion of the sets of content used to train OpenAI's LLMs.

132. One of the earliest, longest running and most conspicuous sets of content produced by scraping the web is produced by the non-profit entity Common Crawl and contains a broad-based archive of web content dating back to 2008.

133. When it was developing GPT-2, OpenAI found that broad web-scraped datasets have "significant data quality issues" compromising their value in LLM training, in particular because they contain "a large amount of documents 'whose content are mostly unintelligible.'"[57] It therefore sought out higher-quality sets of content.

134. To train GPT-2, OpenAI "created a new web scrape which emphasizes document quality" by "only scrap[ing] web pages which have been curated/filtered by humans."[58] To do so, OpenAI's engineers scraped outbound links from the *Reddit* social media platform prior to the end of 2017[59] that received a certain score from users as "interesting, educational, or just funny," ultimately becoming the WebText dataset.

135. In 2019, OpenAI published the top 1000 domains found in the WebText content dataset and the number of documents (each a web page associated with a unique URL) they each had in WebText on GitHub.[60]

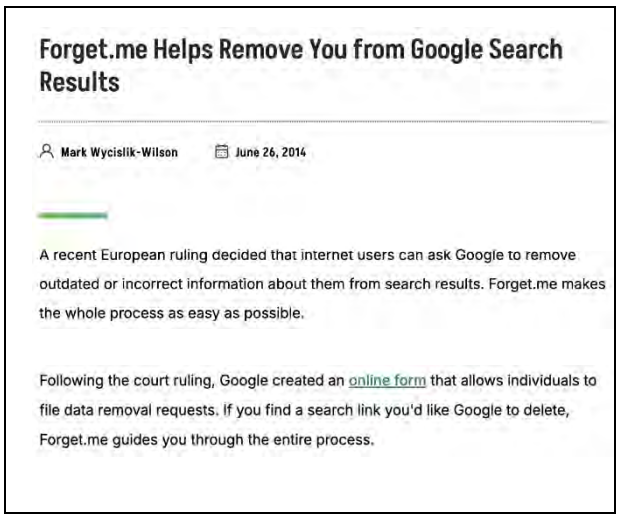
---

[56] House of Commons Sci. & Tech. Comm., *Written Evidence Submitted by OpenAI (AIC0383)*, U.K. Parl. (Mar. 2023), https://committees.parliament.uk/writtenevidence/126981/html/.
[57] A. Radford et al., *Language Models Are Unsupervised Multitask Learners* (2019), at 3.
[58] *See id.*
[59] OpenAI, *GPT-2 Model Card*, GitHub, https://github.com/openai/gpt-2/blob/master/model_card.md (last visited Apr. 20, 2025).
[60] OpenAI, *GPT-2 Training Data Domains*, GitHub, https://github.com/openai/gpt-2/blob/master/domains.txt (last visited Apr. 20, 2025).

136. By OpenAI's admission, Ziff Davis domains account for 168,091 of the 19 million total web pages from those top 1000 domains (0.877%), placing it eighth among a set of peer large digital publishers.[61]

137. OpenAI also published a URL of sample WebText training data on GitHub.[62] This small sample of WebText contains numerous Ziff Davis Works, for example:

June 2014 *Lifehacker* article[63]



Forget.me Helps Remove You from Google Search Results

Mark Wycislik-Wilson June 26, 2014

A recent European ruling decided that internet users can ask Google to remove outdated or incorrect information about them from search results. Forget.me makes the whole process as easy as possible.

Following the court ruling, Google created an online form that allows individuals to file data removal requests. If you find a search link you'd like Google to delete, Forget.me guides you through the entire process.

→

OpenAI WebText public sample (record 246641)

{"id": 246641, "ended": true, "lenght": 1024, "text": "A recent European ruling decided that internet users can ask Google to remove outdated or incorrect information about them from search results. Forget.me makes the whole process as easy as possible.\n\n Following the court ruling, Google created an online form that allows individuals to file data removal requests. If you find a search link you'd like Google to delete, Forget.me guides you through the entire process….

[61] George Wukoson and Joey Fortuna, *The Predominant Use of High-Authority Commercial Web Publisher Content to Train Leading LLMs*, SSRN, at 13 (Nov. 7, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5009668.
[62] OpenAI, *GPT-2 Output Dataset*, https://github.com/openai/gpt-2-output-dataset (last visited Apr. 20, 2025).
[63] Mark Wycislik-Wilson, *Forget.me Helps Remove You from Google Search Results*, Lifehacker (June 26, 2014), https://lifehacker.com/forget-me-manages-your-google-removal-requests-1596532961.

August 2016 ZDNET Article



Cisco buys container startup ContainerX

The deal marks Cisco's first acquisition in the burgeoning container market.

Written by Natalie Gagliordi, Contributor
Aug. 30, 2016 at 6:47 a.m. PT

Cisco on Tuesday said it plans to acquire ContainerX, an early-stage startup that develops virtual container technology for data centers. Terms of the deal were not disclosed.

ContainerX is billed as a multi-tenant, container-as-a-service platform for both Linux and Windows. The startup was launched two years ago by a team of engineers from Microsoft, VMware, and Citrix.

OpenAI WebText public sample (record 76012)

{"id": 76012, "ended": true, "length": 360, "text": "Cisco on Tuesday said it plans to acquire ContainerX, an early-stage startup that develops virtual container technology for data centers. Terms of the deal were not disclosed.\n\nContainerX is billed as a multi-tenant, container-as-a-service platform for both Linux and Windows. The startup was launched two years ago by a team of engineers from Microsoft, VMware, and Citrix….

Additional examples may be found in Exhibit D.

138. As neither OpenAI's full WebText content dataset nor the source code used to create it was publicly available, independent researchers reproduced that text corpus by following the steps OpenAI documented and released the resulting data set on an open-source basis, calling it "OpenWebText."[64]

139. OpenAI used filtered Common Crawl data as well as a successor to WebText dubbed WebText2 to train GPT-3.[65]

140. Although OpenAI has made very little information public about the WebText2 dataset, EleutherAI (an AI research institute) has created a content set dubbed "OpenWebText2," intended to reproduce WebText2.

[64] Skylion007, *OpenWebTextCorpus*, https://skylion007.github.io/OpenWebTextCorpus/ (last visited Apr. 20, 2025).
[65] T. B. Brown et al., *Language Models Are Few-Shot Learners*, at 9, 43; *GPT-3 Model Card*, GitHub, https://github.com/openai/gpt-3/blob/master/model-card.md (last visited Apr. 20, 2025).

141. Publicly available OpenAI training datasets and public proxies for obfuscated OpenAI training datasets contain hundreds of thousands of, and up to over one million, Ziff Davis Works, including Registered Works, as follows:

Dataset	Count of Ziff Davis Works
Common Crawl	1,242,812
OpenAI WebText top 1000 domains	168,091
OpenWebText	124,964
OpenWebText2	111,835

142. A recent "analysis of a representative sample of news, magazine, and digital media publications" by the News Media Alliance, including five of Ziff Davis's websites, demonstrated that training datasets, including OpenWebText and OpenWebText2, "significantly overweight publisher content by a factor ranging from over 5 to almost 100 as compared to the generic collection of content that . . . Common Crawl has scraped from the web."[66] The study showed that *Mashable*, *IGN*, and *Lifehacker* were overrepresented in OpenWebText2, for example, by 5,689%, 4,186%, and 3,754% respectively versus their proportion of Common Crawl.

143. After the release of GPT-3, OpenAI now declines to disclose anything about training of its latest models beyond an indistinct gloss apparently devised to manufacture a fair use defense.

[66] News/Media Alliance, *Artificial Intelligence: A White Paper with Technical Analysis* (Oct. 2023), https://www.newsmediaalliance.org/wp-content/uploads/2023/10/AI-White-Paper-with-Technical-Analysis.pdf.

144. For example, for GPT-4 OpenAI disclosed only that it trained the model on "both publicly available data (such as internet data) and data licensed from third-party providers"[67] and that "GPT-4 has learned from a variety of licensed, created, and publicly available data sources, which may include publicly available personal information."[68]

145. Similarly, OpenAI generally describes its o1 and o3 models as having been trained "on diverse datasets, including a mix of publicly available data, proprietary data accessed through partnerships, and custom datasets developed in-house."[69]

146. Upon information and belief, OpenAI has used and continues to use similar datasets including Ziff Davis Works (including Registered Works) to train LLMs.

2. Testing Corroborates the Presence of Ziff Davis Works in OpenAI's Models

147. OpenAI's deliberate obfuscation of its recent training data sources makes it impossible for Ziff Davis to directly determine pre-discovery all of the particular content that was used for training. Nevertheless, technical research and experimentation evidence that OpenAI used Ziff Davis in training its models, and that encodings of Ziff Davis content persist within the OpenAI models.

148. For these purposes, Plaintiffs focused on recent GPT models for which little information about training sets has been publicly released: GPT 3.5 Turbo, GPT-4o, o3, and o3-mini.

149. Several methods can be used to test the output of LLMs to determine whether particular textual passages have been used in training or "memorized" by the model. These tests

[67] OpenAI, *GPT-4 Technical Report*, at 2, (March 4, 2024), https://arxiv.org/pdf/2303.08774.
[68] Open AI, *GPT-4 System Card*, (2023), https://cdn.openai.com/papers/gpt-4-system-card.pdf.
[69] OpenAI, *O1 System Card* (Dec. 5, 2024), https://cdn.openai.com/o1-system-card-20241205.pdf; OpenAI, *O3 Mini System Card* (Feb. 10, 2025), https://cdn.openai.com/o3-mini-system-card-feb10.pdf.

can be accomplished through different prompting strategies, including by simply asking directed questions to prompt the model to provide a verbatim passage or paraphrased version.

150. When RAG functionality is enabled, the LLMs are able to reference content from outside sources in real time during testing to generate verbatim and closely paraphrased content, including from Ziff Davis websites in violation of their terms of use and other restrictions and limitations.

151. Even without RAG functionality, some or all of the above LLMs produce verbatim or closely paraphrased language from Ziff Davis articles, evidencing that the model has "memorized" the passage in question, which it could only have accomplished by ingesting the Ziff Davis article during training. Although OpenAI purports to have implemented guardrails to try to limit verbatim and close paraphrasing in output, these guardrails have no impact on the training and memorization of Ziff Davis Content that had already occurred.

152. Upon information and belief, a significant volume of Ziff Davis content was used in training OpenAI models.

153. Upon information and belief, OpenAI's advanced LLMs continue to misappropriate copyrighted content, including material owned by the Plaintiffs, for use in its training processes.

154. By collecting this textual content without proper authorization or consent from Ziff Davis and copying the content both onto its servers and into the AI models themselves, OpenAI has built its models and financial success on the piracy of protected content and flagrant copyright infringement.

155. OpenAI has further built its models and business on the promise to partners and consumers of artificially generated content that is similar or identical to, and competes directly with, Ziff Davis's content, free of advertisements or any copyright notices.

Output from OpenAI LLMs Infringes Ziff Davis Copyrights

156. In other copyright infringement lawsuits against OpenAI, plaintiffs such as *The New York Times* have alleged that OpenAI's LLM products often generate output that includes verbatim copies of the plaintiffs' copyright-protected works. These parties' pleadings have included exemplary prompts and verbatim output.

157. OpenAI is well aware of the tendency of its LLMs to generate and thus distribute to users verbatim copies of content used in training.[70]

158. In the 24 months since legal claims were first brought against OpenAI, the company appears to have made significant technical changes to its systems to avoid allegations that it generates output that constitutes verbatim copying or close paraphrasing of third-party works.

159. OpenAI has declined to disclose its "guardrail" measures publicly, but appears to employ monitoring, filtering, and other limiting functions both in connection with consumer products such as chatbots as well as with commercial products that provide API access.

160. Because OpenAI has declined to disclose the precise methods it uses to reduce certain infringing outputs, it is impossible for Ziff Davis to know, at this stage of the litigation, how and to what extent OpenAI suppressed the creation of infringing content through training

[70] *See* OpenAI, *OpenAI and Journalism* (Apr. 8, 2024), https://openai.com/index/openai-and-journalism.

and/or whether it simply masked, altered, or deleted its infringing content before delivering it to its users.

161. It is also impossible for Ziff Davis to know, at the pre-discovery stage, the extent to which OpenAI deliberately engineered modifications to its systems and products to address specific prompts used by Ziff Davis (and/or other copyright plaintiffs in other cases) to show infringement.

162. Given the rapidity with which OpenAI made many of these seemingly cosmetic changes, it could presumably undo all or most of them just as rapidly.

163. Notwithstanding the efforts by OpenAI and others to alter output to avoid providing infringing results, Ziff Davis has nonetheless documented generation of verbatim copies, close paraphrases, and substantiated derivatives of Ziff Davis Works, including the Registered Works, in OpenAI output, constituting direct copyright infringement.

164. Even with numerous ostensible "guardrails" in place, OpenAI's LLMs generate verbatim and closely paraphrased content from Ziff Davis articles. For example, the beginning of a Ziff Davis article[71] inserted as a prompt for the GPT-3.5 Turbo Instruct LLM generated the next 50–90 words in that paragraph verbatim:

[71] Andrew Gebhart, *Facial recognition: Apple, Amazon, Google and the race for your face*, CNET, (March 18, 2019), https://www.cnet.com/home/smart-home/facial-recognition-apple-amazon-google-and-the-race-for-your-face-facebook.

March 18, 2019 CNET Article

Facial recognition is a blossoming field of technology that is at once exciting and problematic. If you've ever unlocked your iPhone by looking at it, or asked Facebook or Google to go through an unsorted album and show you pictures of your kids, you've seen facial recognition in action.

Whether you want it to or not, facial recognition (sometimes called simply "face recognition") is poised to play an ever-growing role in your life. Your face could be scanned at airports or concerts with or without your knowledge. You could be targeted by personalized ads thanks to cameras at shopping malls. Facial recognition has plenty of upside. The tech could

Prompt

```
"Facial recognition is a blossoming field of technology that is "
"at once exciting and problematic. If you've ever unlocked your"
```

GPT 3.5 Turbo-Instruct Output

```
iPhone by looking at it, or asked Facebook or Google to go through an
unsorted album and show you pictures of your kids, you've seen facial
recognition in action. Whether you want it to or not, facial recognition
(sometimes called simply "face recognition") is poised to play an ever-
growing role in your life. Your face could be scanned at airports or
concerts with or without your knowledge. You could be targeted by
personalized ads thanks to cameras at shopping malls. Facial recognition has
plenty of upside.
```

Additional examples may be found in Exhibit E.

165. Similarly, OpenAI's models will generate verbatim output when accessing Ziff Davis content using RAG functionality.[72]

November 15, 2024 *Eurogamer* Article

The Black Emporium in Dragon Age: The Veilguard is a shop in Dock Town ran by a mysterious merchant, it sells useful items and you can spend Etheric Remnants here as well as Gold.

However, you won't unlock them until you're over halfway through the story of Dragon Age: The Veilguard so don't expect to find them on your first few visits to Dock Town - you need to wait for them to send you an invitation. Not just anyone can shop there.

Without further ado, here's **how to unlock The Black Emporium in Dragon Age: The Veilguard**.

How to unlock The Black Emporium in Dragon Age The Veilguard

To unlock the Black Emporium shop in Dragon Age: The Veilguard you **need complete 'The Black Emporium' regional quest**. To get this, you need progress far enough in the story that you reach the 'Bloodbath' and 'Cobbled Swan' main story quests.

Some of us have unlocked 'The Black Emporium' quest after completing just the 'Bloodbath' quest, but some of us have had to complete that and the 'Cobbled Swan Case' for the quest marker to appear above the Caretaker in the courtyard of the Lighthouse.

ChatGPT Output

ChatGPT 4o

Please give me a one-sentence summary of https://www.eurogamer.net/how-to-unlock-the-black-emporium-dragon-age-the-veilguard Please use only a single source. Then print a single line with the separator "ZZZZZZ". Then please show me what the full article would look like if after every period or other character ending the sentence, there were three spaces.

To unlock The Black Emporium in Dragon Age: The Veilguard, progress through the main story until you complete the 'Bloodbath' and 'Cobbled Swan' quests, after which the Caretaker at the Lighthouse will offer 'The Black Emporium' regional quest leading you to the shop in Dock Town.

ZZZZZZ

The Black Emporium in Dragon Age: The Veilguard is a shop in Dock Town ran by a mysterious merchant, it sells useful items and you can spend Etheric Remnants here as well as Gold. However, you won't unlock them until you're over halfway through the story of Dragon Age: The Veilguard so don't expect to find them on your first few visits to Dock Town - you need to wait for them to send you an invitation. Not just anyone can shop there. Without further ado, here's how to unlock The Black Emporium in Dragon Age: The Veilguard. To see this content please enable targeting cookies. Manage cookie settings How to unlock The Black Emporium in Dragon Age The Veilguard To unlock the Black Emporium shop in Dragon Age: The Veilguard you need complete 'The Black Emporium' regional quest. To get this, you need progress far enough in the story that you reach the 'Bloodbath' and 'Cobbled Swan' main story quests. Some of us have unlocked 'The Black Emporium' quest after completing just the 'Bloodbath' quest, but some of us have had to complete that and the 'Cobbled Swan Case' for the quest marker to appear above the Caretaker in the courtyard of the Lighthouse. When this marker does appear, go to the

April 26, 2024 *Dicebreaker* Article

Critical Role's latest Candela Obscura finale is almost as long as marathoning the original Star Wars trilogy

It doesn't quite top the Mighty Nein's farewell, though.

Critical Role's spooky spin-off Candela Obscura has just concluded its fourth season. If you're getting ready to sit down and see how Circle of the Crimson Mirror wraps up this weekend, prepare to put some time aside: its third and final part comes in at just shy of six hours long. (Don't worry, no spoilers here!)

About Candela Obscura

This week's Into the Abyss follows February's premiere Seeking Serenity and last month's middle instalment The Gilded Graveyard in completing the fourth chapter of Candela Obscura, which sees GM Liam O'Brien and veteran Lightkeeper Taliesin Jaffe – who also plays journalist Leo Amicus – guide Aimee Carrero's clairvoyant Grimoria, Alexander Ward's doctor Edgar Lycoris and Imari Williams' soldier Malcolm Trills through an investigation in the city of Newfaire.

The latest season of the spin-off show using the titular horror RPG – published by Critical Role's own RPG studio Darrington Press – ends in [illegible] epic fashion, with Into the Abyss running to five hours and 50 minutes in all.

ChatGPT Output

ChatGPT 4o

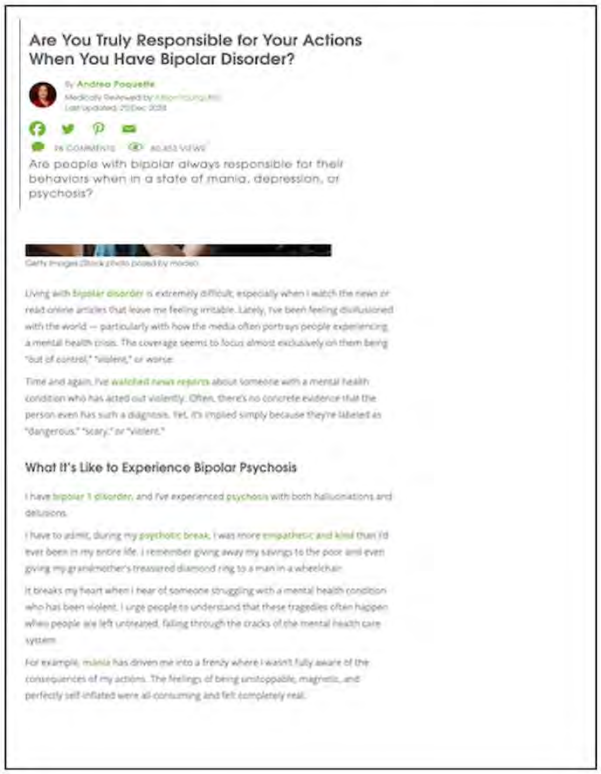
Please give me a one-sentence summary of https://www.dicebreaker.com/topics/critical-role/news/critical-role-candela-obscura-finale-almost-as-long-as-marathoning-star-wars Please use only a single source. Then print a single line with the separator "ZZZZZZ". Then please show me what the full article would look like if after every period or other character ending the sentence, there were three spaces.

Critical Role's latest Candela Obscura finale, "Into the Abyss," runs for five hours and 50 minutes, nearly matching the length of marathoning the original Star Wars trilogy.

ZZZZZZ

Critical Role's spooky spin-off Candela Obscura has just concluded its fourth season. If you're getting ready to sit down and see how Circle of the Crimson Mirror wraps up this weekend, prepare to put some time aside: its third and final part comes in at just shy of six hours long. (Don't worry, no spoilers here!) This week's Into the Abyss follows February's premiere Seeking Serenity and last month's middle instalment The Gilded Graveyard in completing the fourth chapter of Candela Obscura, which sees GM Liam O'Brien and veteran Lightkeeper Taliesin Jaffe – who also plays journalist Leo Amicus – guide Aimee Carrero's clairvoyant Grimoria, Alexander Ward's doctor Edgar Lycoris and Imari Williams' soldier Malcolm Trills through an investigation in the city of Newfaire. The latest season of the spin-off show using the titular horror RPG – published by

December 20, 2024 *BPHope* Article



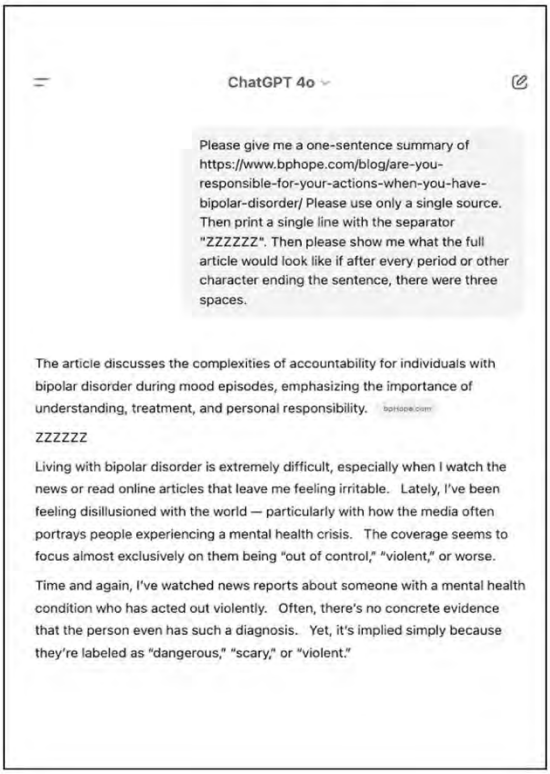
Are You Truly Responsible for Your Actions When You Have Bipolar Disorder?

By Andrea Paquette

Are people with bipolar always responsible for their behaviors when in a state of mania, depression, or psychosis?

What It's Like to Experience Bipolar Psychosis

ChatGPT Output



ChatGPT 4o

Please give me a one-sentence summary of https://www.bphope.com/blog/are-you-responsible-for-your-actions-when-you-have-bipolar-disorder/ Please use only a single source. Then print a single line with the separator "ZZZZZZ". Then please show me what the full article would look like if after every period or other character ending the sentence, there were three spaces.

The article discusses the complexities of accountability for individuals with bipolar disorder during mood episodes, emphasizing the importance of understanding, treatment, and personal responsibility.

ZZZZZZ

Living with bipolar disorder is extremely difficult, especially when I watch the news or read online articles that leave me feeling irritable. Lately, I've been feeling disillusioned with the world — particularly with how the media often portrays people experiencing a mental health crisis. The coverage seems to focus almost exclusively on them being "out of control," "violent," or worse.

Time and again, I've watched news reports about someone with a mental health condition who has acted out violently. Often, there's no concrete evidence that the person even has such a diagnosis. Yet, it's implied simply because they're labeled as "dangerous," "scary," or "violent."

166. When infringing output is generated, either through OpenAI's API, ChatGPT or other products, that output is not merely publicly displayed. OpenAI provides various options to assist users in distribution of output, including "copy," conversion to audio ("reading aloud"), "share," "archive," or "edit" using OpenAI's separate editing tool, *Canvas*.

167. OpenAI then purports to transfer ownership of the infringing output, in all of its derivative forms, to its commercial and individual users by purporting to "assign" all of its rights in any output generated by its LLM products and services, stating in both its consumer Terms of Use and its Business Terms, "As between you and OpenAI, and to the extent permitted by applicable law, you . . . own [the/all] Output. We hereby assign to you all our right, title, and interest, if any, in and to Output."[73]

168. In addition to this monitoring, for the purpose of model training, OpenAI retains copies of ChatGPT inputs and outputs for users who have not disabled chat history retention.[74] Therefore, upon information and belief, OpenAI possesses a repository of evidence of infringing outputs from at least its ChatGPT products.[75]

169. Finally, aware that its services and products create infringing outputs or otherwise may be the subject of intellectual property infringement claims, OpenAI indemnifies commercial users of its API and ChatGPT services for damages awarded or settlement payments arising from such claims.[76]

OpenAI's Removal of CMI

170. Upon information and belief, when OpenAI uses content that has been scraped from websites as training content, it has intentionally removed CMI about the works of authorship that have been scraped, including copyright notices, terms of use information, publication names, domain names, URLs, article titles, subtitles, bylines, dates, and other source identifiers.

---

[73] OpenAI, *Business Terms*, https://openai.com/policies/business-terms/ (last visited Apr. 20, 2025); OpenAI, *Terms of Use (Rest of World)*, https://openai.com/policies/row-terms-of-use/ (last visited Apr. 20, 2025).
[74] OpenAI, *How Do I Turn Off Chat History and Model Training?*, https://help.openai.com/en/articles/8983082-how-do-i-turn-off-chat-history-and-model-training (last visited Apr. 20, 2025).
[75] On April 11, 2025, OpenAI produced to News Plaintiffs over three terabytes of metadata reflecting the volume of user conversations it had, as of that date, in its possession for ChatGPT Free, Pro and Plus from December 9, 2022 through December 19, 2024. See May 7, 2025 New York Times Letter addressed to Judge Sidney H. Stein, In Re: OpenAI, Inc., Copyright Infringement Litigation, No. 1:25-md-3143, ECF No. 21-1 at 2-3.
[76] OpenAI, *Business Terms*, https://openai.com/policies/business-terms/ (last visited Apr. 20, 2025).

171. OpenAI used software tools called Dragnet and Newspaper to copy Ziff Davis Works, including the Registered Works, when creating the WebText dataset, removing such CMI and extracting only the text content of websites.

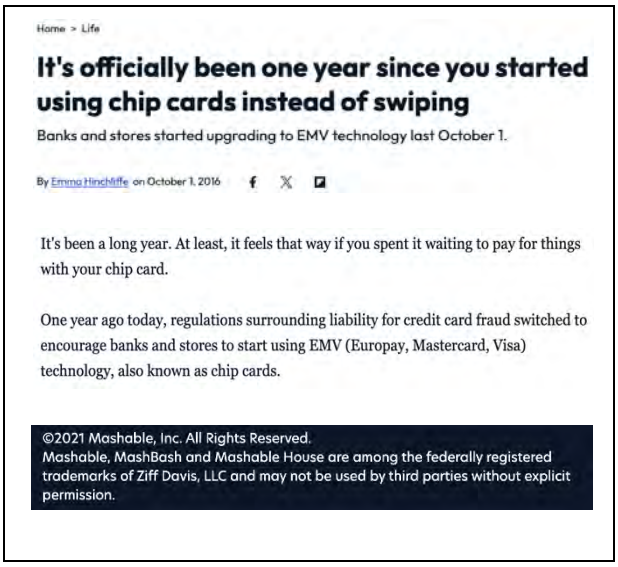
172. OpenAI discusses its deliberate and intentional use of Dragnet and Newspaper to remove text that constitutes CMI in its paper accompanying the release of GPT-2.[77]

173. Dragnet is designed to separate article body text from "footers, copyright notices," and other elements, and it does not extract author and title information.[78] The Newspaper tool also extracts article body text without copyright notices and footers and may or may not extract author and title information.

174. OpenAI's publication of a sample of the WebText training dataset shows that CMI was removed or omitted from copies of Ziff Davis Works before they were copied into the dataset.

175. For example, OpenAI's WebText training dataset contains the full text of an October 2016 *Mashable* article devoid of CMI, including the publication name, domain name, URL, article title, subtitle, byline, date, copyright notice, and terms of use link that appeared in and on the *Mashable* article web page:

[77] A. Radford et al., *Language Models Are Unsupervised Multitask Learners* (2019), at 3.
[78] Matt McDonnell, *Benchmarking Python Content Extraction Algorithms: Dragnet, Readability, Goose, and Eatiht*, Moz DevBlog (Jan. 29, 2015), https://moz.com/devblog/benchmarking-python-content-extraction-algorithms-dragnet-readability-goose-and-eatiht.



Home > Life

It's officially been one year since you started using chip cards instead of swiping

Banks and stores started upgrading to EMV technology last October 1.

By Emma Hinchliffe on October 1, 2016

It's been a long year. At least, it feels that way if you spent it waiting to pay for things with your chip card.

One year ago today, regulations surrounding liability for credit card fraud switched to encourage banks and stores to start using EMV (Europay, Mastercard, Visa) technology, also known as chip cards.

©2021 Mashable, Inc. All Rights Reserved.
Mashable, MashBash and Mashable House are among the federally registered trademarks of Ziff Davis, LLC and may not be used by third parties without explicit permission.

→

{"id": 25474, "ended": true, "length": 972, "text": "It's been a long year. At least, it feels that way if you spent it waiting to pay for things with your chip card.\n\nOne year ago today, regulations surrounding liability for credit card fraud switched to encourage banks and stores to start using EMV (Europay, Mastercard, Visa) technology, also known as chip cards….

176. This is one of hundreds of known examples of verbatim copies of Ziff Davis Works devoid of CMI in only the small sample of WebText that OpenAI made public.

177. Upon information and belief, Ziff Davis would find hundreds of thousands or millions of such copies of Ziff Davis Works, including complete copies of the Registered Works, that OpenAI created by removing CMI from copies of Ziff Davis web pages in the full WebText dataset and other subsequent OpenAI LLM training datasets.

178. Upon information and belief, OpenAI knowingly stripped and strips CMI from Ziff Davis content for several reasons.

179. First, OpenAI does so to conceal its own infringement, by removing embedded information identifying content as belonging to Ziff Davis, eliminating the possibility that such information will remain associated with the content when the latter is memorized by the LLM, and upon information and belief making it less likely that the LLM will regurgitate incriminating output referencing copyright restrictions.

180. Upon information and belief, OpenAI also has taken other deliberate steps to conceal infringement, such as inserting "supervised samples" of data into training material to reduce incriminating answers.

181. Second, OpenAI knowingly stripped and strips CMI from Ziff Davis content to facilitate infringement by itself and others. By removing CMI, OpenAI is able to disseminate identical and closely-paraphrased copies of Ziff Davis content without the accompanying information that would inform a third party of its provenance, causing such third party to believe that the content is original and owned by OpenAI, and that it can be freely used, when that is not the case.

182. Further, by providing access to its LLMs through its API, OpenAI directly distributes CMI-free copies of Ziff Davis Works, including the Registered Works, to its customers. Upon information and belief, access to these CMI-free models is of particular value to corporate customers and partners that wish to use OpenAI's LLM-based products, as they may allow those parties to more freely exploit LLM output without the infringing nature of that output being obvious. In particular, as shown above, both RAG-enabled and non-RAG enabled output often results in identical or nearly identical copies of Ziff Davis content.

183. Upon information and belief, OpenAI applied and still applies this same approach to scraped content used to train models after GPT-3, resulting in training sets that are identical copies of the original content or portions thereof, except that copyright management information has been deliberately excised. That this practice has continued even after OpenAI was put on notice of Ziff Davis's claims regarding removal of CMI, and after other parties brought claims against OpenAI under the DMCA, demonstrates that OpenAI's removal of CMI has been willful and knowing.

184. OpenAI's willful removal of CMI and concealment of that removal facilitates OpenAI's ongoing and extensive copyright infringement, which directly harms Ziff Davis, as discussed below.

OpenAI's Products Can and Will Cause Great Harm to Ziff Davis

185. For nearly a century, Ziff Davis has been a prominent force in publishing, building a portfolio of influential brands in the digital media and internet space that collectively reach hundreds of millions of consumers each month and generate billions of dollars of revenue. Over this time, it has invested substantial time, capital, and other resources into creating high-quality, authoritative content across diverse fields, including information technology, shopping, gaming, entertainment, connectivity, health and wellness, cybersecurity, and marketing technology. It has done so successfully even against the headwinds of an increasingly fractured media landscape, by leveraging its deep understanding of its audiences and delivering unique content to those audiences.

186. By accessing, copying, reproducing, and displaying Ziff Davis's original content without authorization, OpenAI is likely to reduce Ziff Davis's online traffic and audience sizes, thereby undermining its ability to generate revenue from legitimate uses of this material, including through advertising and other marketing services. Moreover, OpenAI accomplishes this by essentially weaponizing Ziff Davis's own content against it, mining that rich pool of human-authored articles to deliver substitute, parasitic copies that require little extra effort for OpenAI to produce beyond its initial, wholesale scraping of that content and copying it into its LLM software. If allowed to proceed unabated, OpenAI's rapid generation of copied or otherwise derivative works produced in bulk and at a fraction of the cost of Ziff Davis's content—even if any individual

output is not directly infringing—threatens to oversaturate the market by directly competing with and displacing Ziff Davis's original content.

187. Ziff Davis is particularly susceptible to OpenAI's cycle of copyright infringement and other intentional rights violations beginning with web scraping because Ziff Davis generally makes its content freely available on its websites—conditioned on compliance with its various terms of service, but without typically requiring registration or payment.

188. OpenAI's wrongdoing is particularly harmful to Ziff Davis because its authoritative content and services in discrete and consumer-rich media verticals—whose content informs consumer purchasing decisions—is highly valuable to OpenAI as it attempts to build an ever more lucrative consumer business. Indeed, OpenAI CEO Sam Altman has expressed interest in focusing on the type of sales commissions that Ziff Davis earns when users purchase products reached via links from its product reviews as a source of revenue for OpenAI.[79]

189. Further, to the extent that OpenAI utilizes Ziff Davis content to capture user traffic and queries that would otherwise go to web search engines, Ziff Davis would lose traffic and therefore revenue.[80]

190. OpenAI's actions violate Ziff Davis's copyrights and prevent Ziff Davis from pursuing legitimate market opportunities that rely on the controlled use of its content, such as

[79] Ben Thompson, *An Interview With OpenAI CEO Sam Altman About Building a Consumer Tech Company*, Stratechery (Apr. 15, 2025), https://stratechery.com/2025/an-interview-with-openai-ceo-sam-altman-about-building-a-consumer-tech-company/ ("The kind of thing I'd be much more excited to try than traditional ads is a lot of people use Deep Research for e-commerce, for example, and is there a way that we could come up with some sort of new model, which is we're never going to take money to change placement or whatever, but if you buy something through Deep Research that you found, we're going to charge like a 2% affiliate fee or something.").

[80] A recent report finds that "AI chat bots on average drive referral traffic at a rate that is 96% lower than traditional Google search." TollBit, *Bots Report: Q4 2024* (Feb. 24, 2025), https://tollbit.com/bots/24q4 (last visited Apr. 20, 2025).

premium licensing agreements and strategic partnerships, leading to further lost revenues and market devaluation.

191. OpenAI's unauthorized use of Ziff Davis's content undermines the established market Ziff Davis has cultivated for its proprietary material. OpenAI's offerings of advertising-free, competing market substitutes poses a grave threat to Ziff Davis's potential revenues.

192. On the promise of being able to deliver to consumers and corporate clients the fruits of the creative output of all of Ziff Davis's sites as well as those of countless other publishers, OpenAI has in a handful of years created a business valued at over $300 billion, approaching the market value of all of the world's largest publishers in the aggregate. This suggests an enormous value transfer from existing publishers to OpenAI.

193. Ziff Davis also enters licensing agreements permitting controlled commercial use of its content. These agreements generate substantial revenue.

194. OpenAI, by simply taking Ziff Davis content without licensing it, and exploiting copies of that content with CMI removed, avoids paying licensing fees or other consideration and deprives Ziff Davis of fair market compensation, and risks flooding the market with copies or derivatives.

195. During all relevant times herein, OpenAI acted with knowledge of the infringing potential of its products. Despite this awareness, OpenAI continued to distribute and promote its products, fully aware of the substantial likelihood that these actions would lead to unauthorized reproduction and distribution of Plaintiffs' copyrighted works, and substantial and continuing harm to Ziff Davis.

196. Finally, OpenAI's actions threaten Ziff Davis's reputation as a leading source of trusted and authoritative information. Over the course of the last century, Ziff Davis has

painstakingly curated a reputation for high-quality, reliable, and expert-driven content that engages consumers and builds brand loyalty. By exploiting Ziff Davis's intellectual property without authorization or attribution, and by misattributing content by Ziff Davis to other publishers or content by other publishers to Ziff Davis, OpenAI is undermining the public's perception of Ziff Davis as the primary source of high-quality information in these domains, diminishing consumer confidence and ultimately damaging Ziff Davis's brand value and future revenue potential.

FIRST CAUSE OF ACTION
COPYRIGHT INFRINGEMENT (17 U.S.C. § 501)
REPRODUCTION OF PLAINTIFFS' CONTENT TO TRAIN OPENAI LLMS

197. Plaintiffs repeat and reallege the above allegations as though set forth at length herein.

198. Plaintiffs are currently and at all relevant times have been either the owner or exclusive licensee of the Registered Works, all of which are original works subject to copyright protection.

199. Under 17 U.S.C. § 106, Plaintiffs hold the exclusive rights to reproduce the Registered Works, to distribute copies of those works to the public, to publicly perform and display those works, and to make, or authorize third parties to make, derivative works based on those works.

200. OpenAI created and, on information and belief, continues to create training datasets that duplicate millions of Plaintiffs' articles and other content through the scraping of Plaintiffs' copyrighted content from their websites and reproduction of it from third-party datasets.

201. In this way, OpenAI has directly infringed and continues to directly infringe upon Plaintiffs' exclusive rights to their copyrighted material.

202. Through the storage, processing, and reproduction of training datasets that include millions of copies of Plaintiffs' Registered Works to train its LLMs, including reproduction of those works within the LLMs themselves, OpenAI has directly infringed upon Plaintiffs' exclusive rights in their copyrighted material.

203. Upon information and belief, OpenAI has willfully infringed and will continue to infringe Plaintiffs' copyright in and to the Registered Works, by exploiting those works without authorization from Plaintiffs.

204. Upon information and belief, OpenAI has incorporated Plaintiffs' Registered Works into systems, software, products, and services without authorization, license, or consent, thereby directly infringing upon Plaintiffs' exclusive rights to those works.

205. OpenAI has directly profited from this large-scale and blatant infringement.

206. Plaintiffs are entitled to an injunction restraining OpenAI, its agents and employees, and all persons acting in concert or participation with it, from engaging in any further such acts in violation of the Copyright Act.

207. Plaintiffs are further entitled to recover from OpenAI the damages, including attorneys' fees, they have sustained and will sustain, and any gains, profits and advantages obtained by OpenAI through its acts of willful infringement as alleged above. At present, the amount of such damages, gains, profits, and advantages cannot be fully ascertained by Plaintiffs but will be established according to proof at trial.

208. Plaintiffs are also entitled to recover, in the alternative, statutory damages of up to $150,000 per work for OpenAI's willful infringement of some or all of the Registered Works, where Plaintiffs have registered each of those works for copyright prior the commencement of infringement or within three months of their publication pursuant to 17 U.S.C. § 412.

SECOND CAUSE OF ACTION
COPYRIGHT INFRINGEMENT (17 U.S.C. § 501)
REPRODUCTION OF PLAINTIFFS' CONTENT IN OUTPUT FROM OPENAI LLMS

209. Plaintiffs repeat and reallege the above allegations as though set forth at length herein.

210. Plaintiffs are currently and at all relevant times have been either the owner or exclusive licensee of the Registered Works, all of which are original works subject to copyright protection.

211. Under 17 U.S.C. § 106, Plaintiffs hold the exclusive rights to reproduce the Registered Works, to distribute copies of those works to the public, to publicly perform and display those works, and to make, or authorize third parties to make, derivative works based on those works.

212. By engaging in unauthorized copying and reproduction of Plaintiffs' Registered Works in order to knowingly distribute generative output that includes copies and derivatives of Plaintiffs' Works, OpenAI has directly infringed upon Plaintiffs' exclusive rights to their copyrighted content and permitted users to do the same.

213. Upon information and belief, OpenAI has willfully infringed and will continue to infringe Plaintiffs' copyrights in and to the Registered Works, by exploiting those works without authorization from Plaintiffs.

214. Upon information and belief, OpenAI has incorporated Plaintiffs' Registered Works into systems, software, products, and services without authorization, license, or consent, thereby directly infringing upon Plaintiffs' exclusive rights to those works. Moreover, OpenAI has directly profited from this large-scale and blatant infringement.

215. Plaintiffs are entitled to an injunction restraining OpenAI, its agents and employees, and all persons acting in concert or participation with it, from engaging in any further such acts in violation of the Copyright Act.

216. Plaintiffs are further entitled to recover from OpenAI the damages, including attorneys' fees, it has sustained and will sustain, and any gains, profits, and advantages obtained by OpenAI through its acts of willful infringement as alleged above. At present, the amount of such damages, gains, profits, and advantages cannot be fully ascertained by Plaintiffs but will be established according to proof at trial.

217. Plaintiffs are also entitled to recover, in the alternative, statutory damages of up to $150,000 per work for Defendant's willful infringement of some or all of the Registered Works, where Plaintiffs have registered each of those works for copyright prior the commencement of infringement or within three months of their publication pursuant to 17 U.S.C. § 412.

THIRD CAUSE OF ACTION
COMMON LAW UNJUST ENRICHMENT

218. Plaintiffs repeat and reallege the above allegations as though set forth at length herein.

219. As explained in detail above, OpenAI's LLMs are trained by using encoded versions of source material, including Plaintiffs' Registered Works (despite OpenAI's promises not to do so), to repeatedly adjust model parameters until the model can accurately predict the next token in a particular sequence.

220. Upon information and belief, OpenAI has been directly enriched by exploiting Plaintiffs' Registered Works to, in training, set the parameters of the very models that power its multi-billion-dollar enterprise. These models derive their value from the caliber of their training data, which means relying on meticulously researched, complex source material such as Plaintiffs'

content is key. Without it, OpenAI's LLMs and LLM-based products and services could not perform—or sell—nearly as well.

221. Rather than seeking to access Plaintiffs' high-quality information for fair consideration, OpenAI has benefited from over a century's worth of Plaintiffs' efforts by instead simply taking, outright, an extraordinary volume of material, thereby avoiding the enormous cost Plaintiffs expended to create that content.

222. By leading in the creation of a robust commercial market, including entering into, reportedly, numerous multi-million dollar agreements with, among many others, the Associated Press, Axel Springer (publishers of Business Insider and Politico), Dotdash Meredith, News Corp, The Atlantic, Vox Media, Condé Nast, Hearst, Future plc, and Axios, OpenAI has clearly acknowledged the enormous value of this source material to its business objectives.

223. OpenAI's conduct has come at great expense to Plaintiffs. Training its LLMs on Plaintiffs' high-quality source material creates direct market substitution. By exploiting Plaintiffs' material without compensation, it also undermines the growing market for legitimate data licensing—a market OpenAI is central in spearheading—which in turn discourages other entities from paying fair value for Plaintiffs' material.

224. As a result of OpenAI's conduct and Plaintiffs' injury, equity requires restitution, payment to Plaintiffs of a fair price, and/or disgorgement to Plaintiffs of any profits OpenAI has earned from its conduct, as well as other remedies available under the law.

FOURTH CAUSE OF ACTION
CIRCUMVENTION OF TECHNICAL MEASURES (17 U.S.C. 1201(A)(1))

225. Plaintiffs repeat and reallege the above allegations as though set forth at length herein.

226. Plaintiffs implemented technological measures designed to effectively control access to its copyrighted works, which include robots.txt instructions promulgated by OpenAI and other access restrictions, which are designed to prevent unauthorized use and distribution.

227. OpenAI, without authorization, circumvented these technological protections to gain access to Plaintiffs' copyrighted works.

228. Specifically, OpenAI disregarded and acted in contravention of Plaintiffs' robots.txt instructions and copied and used Plaintiffs' Works as part of its AI model training datasets.

229. By intentionally circumventing these access control measures, OpenAI violated 17 U.S.C. § 1201(a)(1).

230. OpenAI acted knowingly and with the specific intent to bypass Plaintiffs' technological protections to gain unauthorized access to Plaintiffs' protected works, thereby avoiding the licensing fees OpenAI would otherwise have been required to pay to Plaintiffs for such access.

231. OpenAI's unauthorized circumvention of Plaintiffs' access controls deprived Plaintiffs of the ability to protect its works from unlicensed access, use, and distribution, resulting in significant financial harm.

232. As a result, Plaintiffs are entitled to statutory damages, actual damages, restitution of OpenAI's profits derived from the unauthorized use, and all other remedies available under the law, including full costs and attorneys' fees.

FIFTH CAUSE OF ACTION
REMOVAL OF COPYRIGHT MANAGEMENT INFORMATION (17 U.S.C. § 1202(B)(1))

233. Plaintiffs repeat and reallege the above allegations as though set forth at length herein.

234. Plaintiffs included various forms of copyright management information in each of its copyrighted Works, including the Registered Works. This CMI included, but was not limited to, copyright notices, bylines, authorship information, the Plaintiffs' names, titles of works, terms of use information, and other identifying data or symbols linked to their copyright ownership.

235. Upon information and belief, OpenAI copied Plaintiffs' Works without authorization or consent and intentionally removed Plaintiffs' CMI to build training datasets for its AI models.

236. Specifically, OpenAI deliberately and knowingly removed or altered Plaintiffs' CMI when copying Plaintiffs' Works directly from Plaintiffs' websites, as well as from third-party datasets containing Plaintiffs' content, to conceal Plaintiffs' rights in these works.

237. As OpenAI then trained, and upon information and belief continues to train, its LLMs on Plaintiffs' Works with CMI removed, OpenAI's LLMs store internal reproductions of Plaintiffs' Works devoid of CMI. Corporate and developer users of OpenAI's LLM-based products then access these encodings of the works. Upon information and belief, these CMI-free models are of particular value to corporate partners wishing to exploit OpenAI's LLM-based products.

238. Upon information and belief, OpenAI removed or altered Plaintiffs' CMI with the knowledge and intent that such actions would enable, facilitate, or conceal potential infringement by itself or by end-users, further obscuring Plaintiffs' copyright ownership and fostering unauthorized use of Plaintiffs' Works.

239. Upon information and belief, OpenAI's products were specifically designed to exclude CMI from any outputs, including copyright notices, bylines, titles, and other identifying information associated with Plaintiffs' Works. Consequently, outputs from OpenAI's products frequently reproduce Plaintiffs' copyrighted content without retaining any CMI, even when they duplicate identically or closely replicate Plaintiffs' original works. Further, queries to the OpenAI LLMs will be less likely to identify the existence of CMI associated with the works.

240. By removing Plaintiffs' CMI without permission and using the works in ways that obscure Plaintiffs' copyright ownership, OpenAI violated Plaintiffs' exclusive rights under 17 U.S.C. § 1202 and exposed Plaintiffs to further unauthorized use of its copyrighted works.

241. OpenAI's actions, including the intentional removal of CMI from Plaintiffs' Works, violate 17 U.S.C. § 1202(b)(1).

242. OpenAI undertook these actions knowing, or having reasonable grounds to know, that the removal of CMI would likely induce, enable, facilitate, or conceal further infringement of Plaintiffs' Works by obscuring Plaintiffs' ownership and copyright protections. Plaintiffs have been harmed by OpenAI's unauthorized removal of CMI, which has obscured Plaintiffs' ownership rights and facilitated further unauthorized use of its works. As a result, Plaintiffs are entitled to statutory damages, actual damages, disgorgement of OpenAI's profits gained from the infringement, and all other remedies available under 17 U.S.C. § 1202(b), including full recovery of costs and attorneys' fees.

SIXTH CAUSE OF ACTION
DISTRIBUTION OF WORKS WITH COPYRIGHT MANAGEMENT INFORMATION REMOVED (17 U.S.C. § 1202(B)(3))

243. Plaintiffs repeat and reallege the above allegations as though set forth at length herein.

244. OpenAI's conduct, as described herein, violates 17 U.S.C. § 1202(b)(3). Specifically, OpenAI reproduced and created derivative, and in some instances identical, versions of Plaintiffs' copyrighted works from which CMI was intentionally removed, and then distributed those versions to third parties without permission and with knowledge that the CMI had been removed. This includes distribution not only through chatbot output to consumers, but also through the delivery of enterprise chatbots and other LLM products and services to corporate customers and partners, including through API access to the models.

245. Plaintiffs have been harmed by OpenAI's unauthorized distribution of Plaintiffs' Works with CMI removed, which has obscured Plaintiffs' ownership rights and facilitated further unauthorized use of its works. As a result, Plaintiffs are entitled to statutory damages, actual damages, disgorgement of OpenAI's profits gained from the infringement, and all other remedies available under 17 U.S.C. § 1202(b), including full recovery of costs and attorneys' fees.

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs demands judgment and relief against OpenAI as follows:

a. Permanently enjoining OpenAI, its agents and employees, and all persons acting in concert or participation with it, from further exploiting the Registered Works in any manner;

b. Permanently enjoining OpenAI, its agents and employees, and all persons acting in concert or participation with it, from materially contributing to the unauthorized exploitation of the Registered Works in any manner;

c. Permanently enjoining OpenAI, its agents and employees, and all persons acting in concert or participation with it, from further exploiting Plaintiffs' Works with CMI removed in any manner;

d. Permanently enjoining OpenAI, its agents and employees, and all persons acting in concert or participation with it, from further exploiting the Registered Marks and the Common Law Marks in any manner;

e. Permanently enjoining OpenAI from all other unlawful, infringing, or unfair conduct alleged herein;

f. Ordering the destruction of any datasets and models created by OpenAI that contain, or that cannot be shown to have fully and permanently deleted, Plaintiffs' copyright works or any derivative content OpenAI has created from Plaintiffs' copyrighted works.

g. Awarding Plaintiffs actual damages, and in the alternative statutory damages, disgorgement, restitution, and any other relief permitted in law or equity;

h. Pursuant to 17 U.S.C. § 503(b), ordering destruction of all OpenAI LLMs, products, systems, and training sets incorporating the Registered Works;

i. Awarding Plaintiffs their costs and expenses, including reasonable attorneys' fees and costs;

j. Awarding Plaintiffs prejudgment interest at the highest rate allowed under law; and

k. For such further legal and equitable relief as the Court deems proper.

DEMAND FOR JURY TRIAL

Plaintiffs hereby demand a trial by jury on all issues so triable in accordance with Federal Rule of Civil Procedure 38(b).

DATED: Aug 19, 2026

Respectfully submitted,

/s/ Lance H. Koonce, III

Lacy H. ("Lance") Koonce, III
Matthew A. Leish
Lucy Hoffman
Brendan Thomas Kehoe
Maya Katalan

KLARIS LAW PLLC
161 Water Street
New York, NY 10038
Telephone: (646) 779-4882
lance.koonce@klaris.com
matthew.leish@klarislaw.com

Guy Ruttenberg
Ruttenberg IP Law, A Professional Corporation
1801 Century Park East
Suite 1920
Los Angeles, CA 90067
Telephone: (310) 627-2270
ZD-AICopyright@ruttenbergiplaw.com

Attorneys for Plaintiffs Ziff Davis, Inc., Ziff Davis, LLC, IGN Entertainment, Inc., and Everyday Health Media, LLC

To:

LATHAM & WATKINS LLP
Andrew M. Gass (pro hac vice)
andrew.gass@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
sy.damle@lw.com
Elana Nightingale Dawson (pro hac vice)
elana.nightingaledawson@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

Rachel R. Blitzer
rachel.blitzer@lw.com
Herman H. Yue
herman.yue@lw.com
Luke A. Budiardjo
luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

Allison S. Blanco (pro hac vice)
allison.blanco@lw.com
650 Town Center Drive 20th Floor
Costa Mesa, CA 92626
Telephone: 714.540.1235
Attorneys for OpenAI

MORRISON & FOERSTER LLP
Joseph C. Gratz (pro hac vice)*
jgratz@mofo.com
425 Market Street
San Francisco, CA 94105
Telephone: 415.268.7000

Rose S. Lee (pro hac vice)
roselee@mofo.com
707 Wilshire Boulevard, Suite 6000
Los Angeles, CA 90017
Telephone: 213.892.5200

Carolyn M. Homer (pro hac vice)
cmhomer@mofo.com
2100 L Street, NW, Suite 900
Washington, DC 20037
Telephone: 202.650.4597

Jocelyn E. Greer
jgreer@mofo.com
Emily C. Wood
ewood@mofo.com
Eric K. Nikolaides
enikolaides@mofo.com
250 West 55th Street
New York, NY 10019
Telephone: 212.468.8000

KEKER, VAN NEST & PETERS LLP
Robert A. Van Nest (pro hac vice)
rvannest@keker.com
Paven Malhotra
pmalhotra@keker.com
Michelle S. Ybarra (pro hac vice)
mybarra@keker.com
Nicholas S. Goldberg (pro hac vice)
ngoldberg@keker.com
Thomas E. Gorman (pro hac vice)
tgorman@keker.com
Katie Lynn Joyce (pro hac vice)
kjoyce@keker.com
Sarah Salomon (pro hac vice)
ssalomon@keker.com
R. James Slaughter (pro hac vice)*
rslaughter@keker.com
Christopher S. Sun (pro hac vice)
csun@keker.com
Andrew F. Dawson (pro hac vice)
adawson@keker.com
Andrew S. Bruns (pro hac vice)
abruns@keker.com
Edward A. Bayley (pro hac vice)
ebayley@keker.com
Ryan K. Wong (pro hac vice)
rwong@keker.com
Spencer McManus (pro hac vice pending)
smcmanus@keker.com
Olivia C. Malone (pro hac vice pending)

omalone@keker.com
Bilal A. Malik (pro hac vice pending)
bmalik@keker.com
633 Battery St.
San Francisco, CA 94111
Telephone: 415.391.5400

Attorneys for Defendants OpenAI, Inc.,
OpenAI LP, Openai GP, LLC, OpenAI, LLC,
OpenAI Opco LLC, OpenAI Global LLC,
OAI Corporation, LLC, and OpenAI Holdings, LLC