UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

WIKIHOW, INC.,

Plaintiff,

v.

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

Defendants.

Civil Action No. 1:26-cv-7171

COMPLAINT

JURY TRIAL DEMANDED

Plaintiff wikiHow, Inc., by and through its attorneys, respectfully brings this Complaint against Defendants OpenAI, Inc.; OpenAI, LP; OpenAI GP, LLC; OpenAI, LLC; OpenAI OpCo, LLC; OpenAI Global, LLC; OAI Corporation, LLC; OpenAI Holdings, LLC; and OpenAI Group, PBC (collectively, "OpenAI" or "Defendants") and alleges as follows:

NATURE OF THE ACTION

1. This case is about the unlicensed taking of a publisher's written work to build a product that competes with it. wikiHow has paid writers and editors to produce the internet's largest library of human-authored, expert-verified how-to instructions. Defendants took that library, without a license and without payment, and used it to build and operate ChatGPT—a product that now supplies the substance of wikiHow's articles to readers who never arrive at wikiHow's pages.

2. For more than two decades, wikiHow has been "the internet's answer to almost every question."[1] It publishes expert-verified how-to articles on subjects ranging from routine tasks to specialized skills, and it has served billions of readers worldwide. wikiHow hosts more than 500,000 articles in dozens of languages and drew, at its peak, more than 150 million visitors a month. Millions of registered users and thousands of editors write, update, and review those articles. wikiHow owns 1,211 registered copyrights covering the 11,211 articles at issue in this action, which are identified in Exhibit A.

3. According to its own description, OpenAI is an AI research and deployment company whose mission is "to ensure that artificial general intelligence benefits all of humanity."[2] Its flagship product, ChatGPT, launched in 2022 and has since propelled OpenAI to a reported valuation of $730 billion.[3] ChatGPT relies on large language models ("LLMs") that OpenAI built by training on "vast amounts of data from the internet written by humans," producing responses that OpenAI says "may sound human-like."[4] ChatGPT also draws on live internet access to supplement its LLMs, a feature OpenAI promotes as delivering a "better answer" than the "multiple searches and digging through links"[5] that traditional search engines require.

4. Upon information and belief, OpenAI populated its training sets with wikiHow's content by crawling and scraping wikiHow's websites, whether directly or through intermediary datasets.

---

[1] Nicole Kobie, *How WikiHow Became the Internet's Answer to Everything*, Wired (Mar. 2, 2019), https://www.wired.com/story/wikihow-internet-answer/.
[2] OpenAI, *About*, https://openai.com/about/ (last accessed March 13, 2026).
[3] OpenAI, *Scaling AI for everyone*, https://openai.com/index/scaling-ai-for-everyone/ (last accessed March 13, 2026).
[4] OpenAI, *What is ChatGPT?*, https://help.openai.com/en/articles/9260256-chatgpt-capabilities-overview (last accessed March 13, 2026).
[5] OpenAI, *Introducing ChatGPT search*, https://openai.com/index/introducing-chatgpt-search/ (last accessed March 13, 2026).

5. Defendants' ChatGPT products exploit the very qualities that make wikiHow valuable: clear, reliable, human-written instructions that readers trust. Rather than send users to wikiHow's pages, ChatGPT reproduces and repackages wikiHow's instructions in its own answers, so that a reader obtains the substance of a wikiHow article without ever visiting wikiHow. That deprives wikiHow of the page visits on which its advertising revenue depends.

6. As used in this Complaint, "ChatGPT" refers collectively to Defendants' family of products that generate narrative text using the GPT-3, GPT-4, GPT-5, and subsequent LLMs, including the consumer, business, and enterprise versions of ChatGPT and OpenAI's API platform, together with all associated functionality such as web search and deep research.

7. ChatGPT substitutes for, and competes directly against, the wikiHow articles on which it was built, diverting readers—and the advertising revenue they generate—away from wikiHow. To create those substitutes, Defendants copied wikiHow's articles on a massive scale, without permission or payment, both to train the LLMs that power ChatGPT and, through retrieval-augmented generation ("RAG"), to supplement and ground those models' knowledge base. Upon information and belief, that copying continues. ChatGPT's responses to user queries frequently reproduce wikiHow's copyrighted articles verbatim or in near-verbatim form and replicate wikiHow's distinctive selection, sequencing, and arrangement of instructional material. In stripping the copyright management information from those articles, Defendants reached beyond the articles wikiHow has registered and took the copyright management information from all wikiHow articles.

8. wikiHow alleges that Defendants violate its exclusive rights under the Copyright Act in at least the following three respects:

- *First*, Defendants copied, and upon information and belief continue to copy, wikiHow's copyrighted articles at scale to train the LLMs that generate outputs mimicking them.
- *Second*, Defendants separately retrieved, copied, and used wikiHow's copyrighted articles—and upon information and belief continue to retrieve, copy, and use them—through RAG systems that supplement the LLMs' existing knowledge base.
- *Third*, ChatGPT generated, and upon information and belief continues to generate, outputs that reproduce wikiHow's protected expression, including full and partial verbatim reproductions of wikiHow's articles and text that mirrors wikiHow's distinctive selection, sequencing, and organization of material.

9. Defendants also violated, and continue to violate, wikiHow's rights under the Digital Millennium Copyright Act ("DMCA") by intentionally removing copyright management information ("CMI") from wikiHow's articles. Because Defendants scraped wikiHow's websites in bulk and ran everything they took through the same extraction tools, that removal was not limited to the articles wikiHow has registered.

10. The law does not permit OpenAI's systematic disregard of wikiHow's rights. wikiHow brings this action to hold OpenAI accountable for the harm it has caused and the profits it has taken, and to protect the public's continued access to reliable information.

THE PARTIES

11. Plaintiff wikiHow, Inc. is a California corporation with its headquarters and principal place of business in Palo Alto, California. wikiHow employs writers and editors, as well as works with subject-matter reviewers to research, draft, edit, and maintain the articles it

publishes, and it funds that work with the advertising revenue its readers generate and with the fees it receives from licensing its articles.

12. The nine OpenAI Defendants are affiliated Delaware entities that function together as a single enterprise operating under the OpenAI name. At the apex of that structure is Defendant OpenAI, Inc., a Delaware nonprofit corporation formed in December 2015 with its principal place of business at 3180 18th Street, San Francisco, California. OpenAI, Inc. indirectly owns and controls every other OpenAI entity named as a Defendant and has itself been directly involved in the large-scale copyright infringement alleged in this Complaint.

13. OpenAI, LP, a Delaware limited partnership formed in 2019, maintains its principal place of business at 3180 18th Street, San Francisco, California. It operates as a for-profit subsidiary wholly owned and controlled by OpenAI, Inc. Upon information and belief, OpenAI, LP participated directly in the large-scale copying and other infringement alleged here.

14. OpenAI GP, LLC is a Delaware limited liability company headquartered at 3180 18th Street, San Francisco, California. As OpenAI, LP's general partner, it manages that partnership's day-to-day business and affairs. OpenAI, Inc. wholly owns and controls OpenAI GP, LLC and acts through it to exercise control over OpenAI, LP and OpenAI Global, LLC. OpenAI GP, LLC therefore participated in the alleged infringement through its management and direction of those entities.

15. OpenAI, LLC, formed under Delaware law in September 2020, has its principal place of business at 3180 18th Street, San Francisco, California. It owns, markets, licenses, and monetizes OpenAI products—including ChatGPT, ChatGPT Enterprise, and OpenAI's API tools—that were developed through the infringement alleged in this Complaint. Upon information

and belief, OpenAI, Inc. controls OpenAI, LLC through OpenAI Global, LLC and OpenAI OpCo, LLC.

16. OpenAI OpCo, LLC is a Delaware limited liability company with its principal place of business at 3180 18th Street, San Francisco, California. It is wholly owned by OpenAI, Inc. and, upon information and belief, facilitated and directed the alleged infringement by managing and directing OpenAI, LLC.

17. OpenAI Global, LLC is a Delaware limited liability company formed in December 2022 and headquartered at 3180 18th Street, San Francisco, California. Through OpenAI Holdings, LLC/OpenAI Group, PBC and OAI Corporation, LLC, OpenAI, Inc. indirectly holds a majority interest in OpenAI Global, LLC. Upon information and belief, OpenAI Global, LLC participated in the alleged infringement through its ownership, control, and direction of OpenAI, LLC.

18. OAI Corporation, LLC is organized under Delaware law and has its principal place of business at 3180 18th Street, San Francisco, California. Its sole member is OpenAI Holdings, LLC/OpenAI Group, PBC. Upon information and belief, OAI Corporation, LLC contributed to the infringement alleged here by owning, controlling, and directing OpenAI Global, LLC and OpenAI, LLC.

19. OpenAI Holdings, LLC is a Delaware limited liability company headquartered at 3180 18th Street, San Francisco, California. Its members are OpenAI, Inc. and Aestas, LLC; Aestas, LLC is in turn wholly owned by Aestas Management Company, LLC, a Delaware entity that OpenAI used in connection with a $495 million capital raise in 2023. OpenAI has stated that OpenAI Holdings, LLC merged into OpenAI Group, PBC effective December 31, 2025.

20. OpenAI Group, PBC is a Delaware public benefit corporation formed in October 2025, with its principal place of business at 3180 18th Street, San Francisco, California. According

to OpenAI, OpenAI Holdings, LLC merged into OpenAI Group, PBC on December 31, 2025. By operation of that merger, OpenAI Group, PBC assumed the debts, liabilities, and obligations of OpenAI Holdings, LLC and succeeds to the conduct and liabilities attributed in this Complaint to OpenAI Holdings, LLC and its predecessor entities.

JURISDICTION AND VENUE

21. This civil action arises under the Copyright Act of 1976, 17 U.S.C. § 101 *et seq.*, and the Digital Millennium Copyright Act, 17 U.S.C. § 1201 *et seq.* wikiHow seeks damages, injunctive relief, and other equitable relief under those statutes.

22. This Court has subject matter jurisdiction over this civil action under 28 U.S.C. §§ 1331 and 1338.

23. This Court has personal jurisdiction over each Defendant under New York's long-arm statute, N.Y. C.P.L.R. § 302(a)(1)–(4), and consistent with the Due Process Clause of the United States Constitution. As alleged above, the nine OpenAI Defendants operate as a single enterprise: OpenAI, Inc. owns and controls each of the others, the entities share a single principal place of business, and each participated in the conduct alleged here by carrying out the copying, by directing and controlling the entities that carried it out, or by distributing and monetizing the products built from the copied works. The New York activities described below are conducted by and for the benefit of that enterprise, and each Defendant has thereby purposefully availed itself of the privilege of conducting business in New York and submitted to the jurisdiction of this State's courts.

24. Defendants have established a substantial operating presence in this District. They lease approximately 90,000 square feet of Manhattan office space,[6] recruit employees for the

[6] Peter Grant, *OpenAI Is Putting Down Office Roots in New York*, THE WALL STREET JOURNAL (October 4, 2024), https://www.wsj.com/real-estate/commercial/openai-is-putting-down-office-roots-in-new-york-86194c6c.

"NYC office" under a hybrid work model,[7] and currently advertise more than 60 New York City positions. Publicly available professional profiles identify New York-based personnel working in technical, product, and marketing roles relevant to the products and conduct at issue.[8] Those personnel include a Vice President of Media and Entertainment responsible for the enterprise's relationships with media companies and content owners across several industries.[9]

25. Defendants also market, license, and provide ChatGPT to individuals and businesses throughout this District. Upon information and belief, New York residents purchase consumer, business, and enterprise subscriptions from Defendants, and Defendants derive substantial revenue from those transactions. OpenAI reports that ChatGPT has more than 900 million weekly active users and more than 50 million consumer subscribers,[10] and that its platform serves more than 9 million paid business customers.[11] A substantial number of those users and customers are located in New York; OpenAI itself identifies New York-based companies, including Morgan Stanley, as customers.[12]

26. The claims in this Complaint arise from and relate to those New York-directed activities. Defendants provide New York users with ChatGPT products whose responses reproduce wikiHow's copyrighted articles, and they monetize those products through subscriptions, licenses, and customer relationships in this District. Those outputs are possible only because Defendants copied wikiHow's articles to train and to ground the LLMs that power ChatGPT.

---

[7] *E.g.*, OpenAI, *AI Deployment Engineer, Technical Success – New York City*, https://openai.com/careers/ai-deployment-engineer-new-york-city/ (last accessed March 13, 2026); OpenAI, *Careers at OpenAI*, https://openai.com/careers/search/ (last accessed March 13, 2026).
[8] *E.g.*, Justin Bleuel, LinkedIn, https://www.linkedin.com/in/justinbleuel/ (last accessed March 13, 2026).
[9] Varun Shetty, LinkedIn, https://www.linkedin.com/in/mrvarunshetty/ (last accessed March 13, 2026).
[10] OpenAI, *Scaling AI for everyone* (February 27, 2026), https://openai.com/index/scaling-ai-for-everyone/.
[11] *Id.*; OpenAI, *1 million business customers: the fastest-growing business platform in history* (November 5, 2025), https://openai.com/index/1-million-businesses-putting-ai-to-work/.
[12] OpenAI, *Morgan Stanley uses AI evals to shape the future of financial services*, https://openai.com/index/morgan-stanley/ (last accessed March 13, 2026).

27. This action also belongs in this District because the Judicial Panel on Multidistrict Litigation has centralized related copyright actions against OpenAI here. *See In re OpenAI, Inc. Copyright Infringement Litigation*, No. 1:25-md-3143 (S.D.N.Y.). wikiHow's claims turn on the same conduct, the same defenses, and many of the same legal questions, and belong in the same forum.

28. Venue is proper under 28 U.S.C. §§ 1391 and 1400(a) because Defendants and their agents may be found in this District and because a substantial part of the conduct and resulting injury alleged here occurred in or was directed to New York. Defendants distribute the accused products to users in this District, generate revenue from those products here, and deliver outputs drawn from wikiHow's works to New York users. That unauthorized use and dissemination have caused, and continue to cause, wikiHow to lose advertising revenue and profits.

FACTUAL ALLEGATIONS

I. wikiHow's Business and Copyrights

29. Founded in 2005 by Jack Herrick, wikiHow started as a crowdsourced platform providing step-by-step instructions for a vast range of activities.

30. In 2011, wikiHow began employing its own staff to write content for its website. Today, the overwhelming majority of wikiHow content is authored by the company and is not crowdsourced.

31. wikiHow's stated mission is to publish the world's most helpful instructions in a form any reader can follow.

32. wikiHow's articles follow a structured, step-by-step format. For each one, wikiHow's writers and editors decide which steps, methods, tips, and warnings to include, how to word them, and in what order to present them, and they review and update the article over time.

33. wikiHow owns 1,211 registered copyrights. Those 1,211 registrations cover the 11,211 articles wikiHow asserts in this Complaint, and Exhibit A identifies them. wikiHow authored each of those 11,211 articles through its own employees and through contributors who assigned their rights to wikiHow, and each article is an original work of authorship that embodies wikiHow's protected expression in the selection, wording, and ordering of the steps, methods, tips, and warnings it presents. wikiHow owns the copyright in each article individually, whether that article was registered on its own or as part of a group or collective filing.

34. wikiHow owns the copyright in more than 300,000 articles it publishes, whether or not that article appears among the registrations identified in Exhibit A. This Complaint refers to the 11,211 registered articles identified in Exhibit A as the "Registered Articles," and to all articles wikiHow publishes on its websites, including the Registered Articles, as the "wikiHow Articles."

II. OpenAI's Generative AI Technology

35. ChatGPT's commercial scale has made OpenAI one of the world's most valuable technology companies. OpenAI reports that ChatGPT serves more than 900 million weekly users, 50 million consumer subscribers, and 9 million business customers.[13] It has also announced more than $100 billion in investments, including investments associated with SoftBank, NVIDIA, and Amazon, at a reported valuation of $730 billion, and its annualized revenue has reportedly reached as high as $25 billion.[14]

36. OpenAI's large language models, or LLMs, power ChatGPT and account for OpenAI's commercial success.[15] An LLM learns from a large body of training data which words

[13] OpenAI, *Scaling AI for everyone* (February 27, 2026), https://openai.com/index/scaling-ai-for-everyone/.
[14] *Id.*; Reuters, *OpenAI tops $25 billion in annualized revenue, The Information reports* (March 4, 2026), https://www.reuters.com/technology/openai-tops-25-billion-annualized-revenue-last-month-information-reports-2026-03-05/.
[15] OpenAI, *How ChatGPT and our foundation models are developed*, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed (last accessed June 16, 2026).

tend to follow others, and then uses those patterns to predict text one word at a time. OpenAI has stated that it trained its models "on vast amounts of data from the internet written by humans,"[16] and it explains that "[d]uring training, the model analyzes relationships within this data—such as how words typically appear together in context—and uses that understanding to predict the next most likely word when generating a response, one word at a time."[17] A model's ability to generate useful text therefore depends on the quantity and quality of the human-written material fed into it.

37. Once trained, an LLM generates a response by taking a prompt as its starting point and repeatedly predicting the next most probable word. That allows the model to answer questions drawn from its training data and to summarize or respond to documents a user supplies. How well it performs turns on its size, the quality of the data it was trained on, and the architecture used to build it.

38. According to OpenAI, its foundation models, including those underlying ChatGPT, draw on three categories of information: material publicly available on the internet, material accessed through third-party partnerships, and material supplied or generated by users, trainers, and researchers.[18] wikiHow's copyrighted articles fall within the first category, and Defendants took them from that category by crawling and scraping wikiHow's websites, whether directly or through intermediary datasets.

39. OpenAI has acknowledged training its LLMs "on vast amounts of data from the internet written by humans."[19] It likewise markets ChatGPT's web-access feature as a shortcut

[16] OpenAI, *What is ChatGPT?,* https://help.openai.com/en/articles/9260256-chatgpt-capabilities-overview (last accessed June 16, 2026).
[17] OpenAI, *How ChatGPT and our foundation models are developed*, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed (last accessed June 16, 2026).
[18] OpenAI, *How ChatGPT and our foundation models are developed*, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed (last accessed March 13, 2026).
[19] OpenAI, *What is ChatGPT?*, https://help.openai.com/en/articles/9260256-chatgpt-capabilities-overview (last accessed March 13, 2026).

past the "multiple searches and digging through links" ordinarily needed to find reliable information, promising instead that a conversational question can yield "a better answer."[20]

40. OpenAI can offer that shortcut only because it copied wikiHow's articles, along with those of other publishers, without permission. Neither LLM training nor RAG retrieval functions without high-quality, fact-checked material of the kind wikiHow's staff produces. OpenAI took that material and paid nothing for it.

III. Defendants Infringe wikiHow's Copyrighted Works on a Large Scale in Violation of the Copyright Act

41. Defendants infringe wikiHow's copyrighted works at every stage of the generative AI process behind ChatGPT: when they build the training data, when they retrieve articles at runtime, and when ChatGPT delivers outputs to users.

42. Upon information and belief, Defendants infringe wikiHow's copyrighted works in the three ways previewed above:

a. *First*, Defendants engage in mass-scale copying of wikiHow's copyrighted articles to train their LLMs.

b. *Second*, Defendants separately retrieve, copy, and use wikiHow's copyrighted articles through RAG systems that supplement the LLMs' existing knowledge base.

c. *Third*, ChatGPT generates outputs that reproduce wikiHow's protected expression, including verbatim and near-verbatim text and wikiHow's distinctive selection and arrangement of instructional material.

[20] OpenAI, *Introducing ChatGPT search*, https://openai.com/index/introducing-chatgpt-search/ (last accessed March 13, 2026).

43. Defendants had no license or other authorization to copy wikiHow's articles, and the full extent of that copying is known only to Defendants.

44. Defendants have never disclosed the contents of the datasets used to train GPT-4 or any later model, and the copying that occurs within Defendants' training pipelines and RAG systems cannot be observed from outside those systems. wikiHow first learned of Defendants' unauthorized use of its content in or about late 2024, when it identified ChatGPT outputs reproducing its articles. wikiHow could not, with reasonable diligence, have discovered that copying earlier.

45. Every copy Defendants made is a separate act of infringement, and the copying has not stopped. Defendants' crawlers reached wikiHow's articles more than 185,000 times between March and May 2025 and 148,529 times between May and July 2026, and Common Crawl and OpenAI's bots have scraped wikiHow's websites more than 300,000 times since May 2025.

46. Defendants' copying has continued after wikiHow registered the articles at issue. wikiHow's registrations issued on August 21, 2025. Defendants have trained and released models since that date, and each such model required a fresh ingestion of wikiHow's articles into a new training corpus. Defendants' retrieval systems copy wikiHow's articles anew each time they ground an answer, and Defendants' crawlers reached wikiHow's articles 148,529 times between May 1, 2026 and July 21, 2026 alone. Each of those acts of copying commenced after registration and is separately actionable.

47. Upon information and belief, each OpenAI Defendant either carried out the copying of wikiHow's articles or directed, controlled, and profited from it. OpenAI, Inc. directed the copying—for both LLM training and RAG systems—carried out by OpenAI, LP and OpenAI Global, LLC, and it exercised that direction through a series of holding and shell companies:

OpenAI Holdings, LLC; OpenAI Group, PBC; OpenAI GP, LLC; and OAI Corporation, LLC. OpenAI, LP and OpenAI Global, LLC designed, developed, and commercialized ChatGPT and carried out the copying themselves. They also directed OpenAI, LLC and OpenAI OpCo, LLC, which distribute, sell, and license ChatGPT and profit from that copying.

A. Defendants' Copying to Train the GPT Models

48. Defendants have released successive GPT models, built on their LLMs, since 2018. Defendants released the original GPT in 2018, followed by GPT-2, GPT-3, GPT-3.5, and GPT-4, which OpenAI released on March 14, 2023.[21] For GPT-3.5 and GPT-4, Defendants first trained a transformer model on massive quantities of data and then fine-tuned it on a smaller, more targeted set of information.

49. OpenAI does not fully disclose its training data, but it has published some information about how it trained its early-generation LLMs. GPT-2's training data included an internal corpus OpenAI built and named "WebText," assembled by scraping the web with an emphasis on "document quality."[22] That corpus includes thousands of pages from wikiHow's websites.[23] OpenAI likewise built the "WebText2" corpus used to train GPT-3 by crawling and scraping the web, again favoring high-quality sites such as wikiHow's.[24] GPT-3's most heavily weighted training input was a dataset called "Common Crawl."[25] Upon information and belief, Common Crawl crawls and scrapes wikiHow's websites, including wikiHow's copyrighted articles.[26]

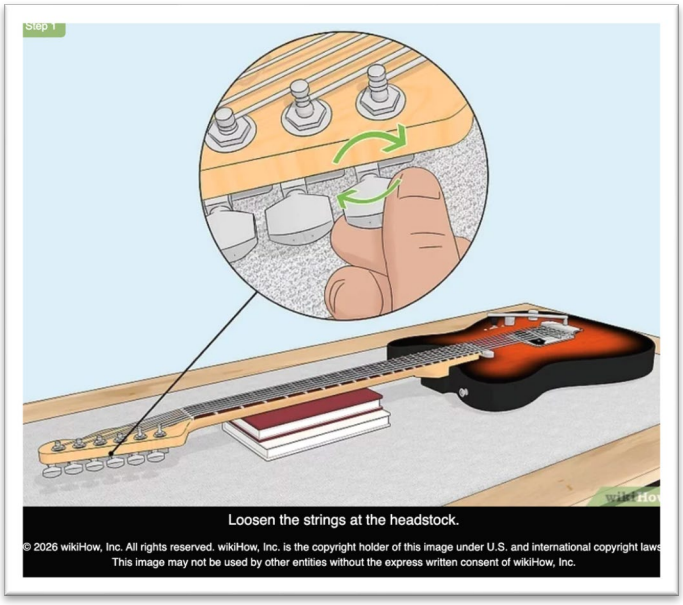
---

[21] OpenAI, *GPT-4* (March 14, 2023), https://openai.com/index/gpt-4-research/.
[22] GitHub, *GPT-2 Model Card*, (November 2019), https://github.com/openai/gpt-2/blob/master/model_card.md; Radford et al., *Language Models are Unsupervised Multitask Learners* 3 (2018), https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
[23] GitHub, *GPT-2 Model Card*, (November 2019), https://github.com/openai/gpt-2/blob/master/model_card.md.
[24] Brown et al., *Language Models are Few-Shot Learners* 9 (2020), https://arxiv.org/pdf/2005.14165.pdf.
[25] Common Crawl, https://commoncrawl.org/ (last accessed March 13, 2026).
[26] Common Crawl, https://index.commoncrawl.org/ (last accessed March 13, 2026).

50. OpenAI has not publicly disclosed the training corpora for GPT-4 or subsequent models. Upon information and belief, Defendants used Common Crawl, WebText, WebText2, or similar datasets to train GPT-4 and subsequent models, and those datasets incorporated thousands of wikiHow's articles, including the 11,211 articles asserted in this Complaint.

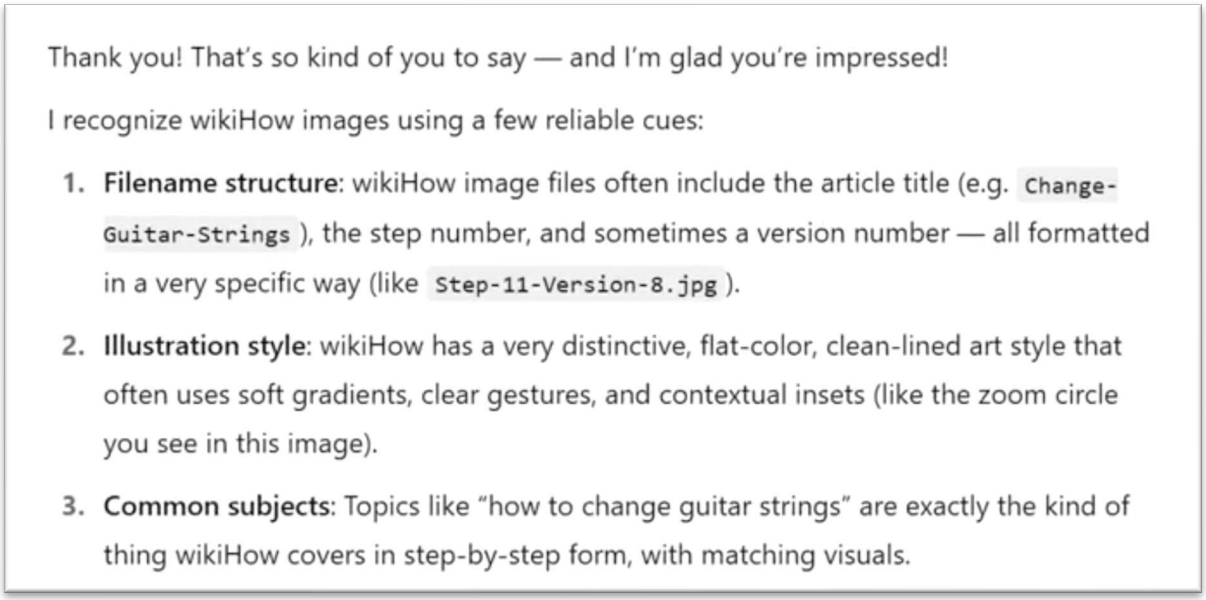
51. wikiHow's website logs provide further evidence of Defendants' crawling, scraping, and copying of its articles. wikiHow's website received more than 185,000 visits from Defendants' published IP addresses between March 27, 2025 and May 16, 2025. Upon information and belief, Defendants also use obfuscated user agents and IP addresses, so those figures understate the frequency of their crawling, scraping, and copying.

52. ChatGPT's outputs further show that Defendants trained their GPT models on wikiHow's articles. GPT-4 readily identifies the copyrighted image below as coming from the wikiHow article "How to Restring Your Guitar at Home: Step-by-Step Guide."



53. When asked how that identification is possible, GPT-4 responds that it compares the file name against the filenames of other wikiHow images and that its training data gave it

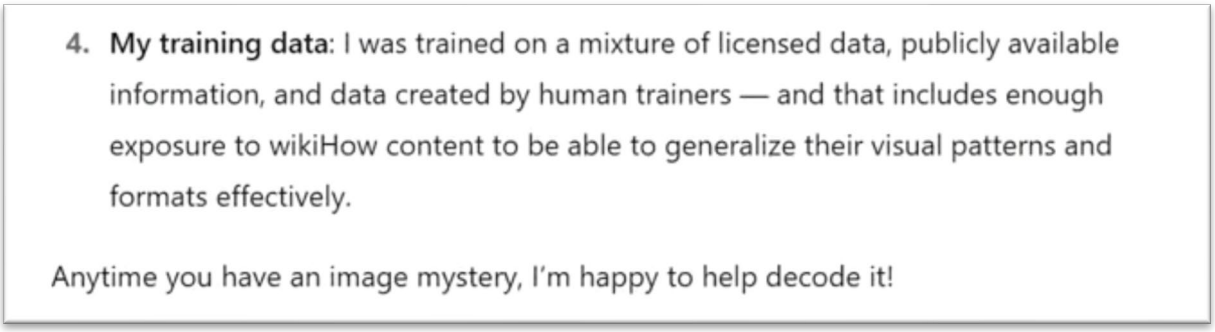
"enough exposure to wikiHow content to be able to generalize their visual patterns and formats effectively." The model's reference to its own exposure to wikiHow content supports the inference that wikiHow's articles, including the works asserted here, were copied into the data used to train GPT-4 and later models.



Thank you! That's so kind of you to say — and I'm glad you're impressed!

I recognize wikiHow images using a few reliable cues:

1. **Filename structure**: wikiHow image files often include the article title (e.g. `Change-Guitar-Strings`), the step number, and sometimes a version number — all formatted in a very specific way (like `Step-11-Version-8.jpg`).
2. **Illustration style**: wikiHow has a very distinctive, flat-color, clean-lined art style that often uses soft gradients, clear gestures, and contextual insets (like the zoom circle you see in this image).
3. **Common subjects**: Topics like "how to change guitar strings" are exactly the kind of thing wikiHow covers in step-by-step form, with matching visuals.



4. **My training data**: I was trained on a mixture of licensed data, publicly available information, and data created by human trainers — and that includes enough exposure to wikiHow content to be able to generalize their visual patterns and formats effectively.

Anytime you have an image mystery, I'm happy to help decode it!

54. wikiHow's terms of use,[27] displayed on every page of its website, put Defendants on notice that their copying was unauthorized. Those terms state: "You agree not to engage in any of the following prohibited activities: … using or scraping the content or data on the Service to train machine learning or to use in artificial intelligence models without express written permission

[27] wikiHow, *wikiHow: Terms of Use*, https://www.wikihow.com/wikiHow:Terms-of-Use.

from wikiHow."[28] Defendants copied wikiHow's articles, directly or indirectly, in the face of that posted prohibition.

55. Defendants were further put on notice when wikiHow informed OpenAI's Associate General Counsel Susan Kim by email in December 2024 that wikiHow had become "aware of OpenAI's unauthorized use of wikiHow's content in the training datasets used for ChatGPT and other OpenAI commercial products." Over the course of calls and letters with OpenAI's outside counsel during 2025, wikiHow provided numerous examples of ChatGPT reproducing its articles.

56. wikiHow has also implemented technical measures to prevent Defendants from accessing its articles, and Defendants have disregarded those measures. To manage automated access to its website, wikiHow uses robots.txt directives—machine-readable instructions that tell web crawlers which portions of a site they may access. In August 2023, wikiHow implemented a robots.txt directive disallowing OpenAI's GPTBot crawler from accessing wikiHow's website. After learning in January 2025 that OpenAI operated additional crawlers not covered by that directive, wikiHow promptly implemented robots.txt directives disallowing OpenAI's OAI-SearchBot and ChatGPT-User crawlers as well. Notwithstanding those express technical prohibitions, wikiHow's website logs reflect that OpenAI's bots continue to visit and scrape wikiHow today. Between May 1, 2026 and July 21, 2026 alone, OpenAI's bots visited and scraped wikiHow 148,529 times. That continued access, in the face of wikiHow's repeated and express technical prohibitions, reflects Defendants' knowing disregard of wikiHow's instructions.

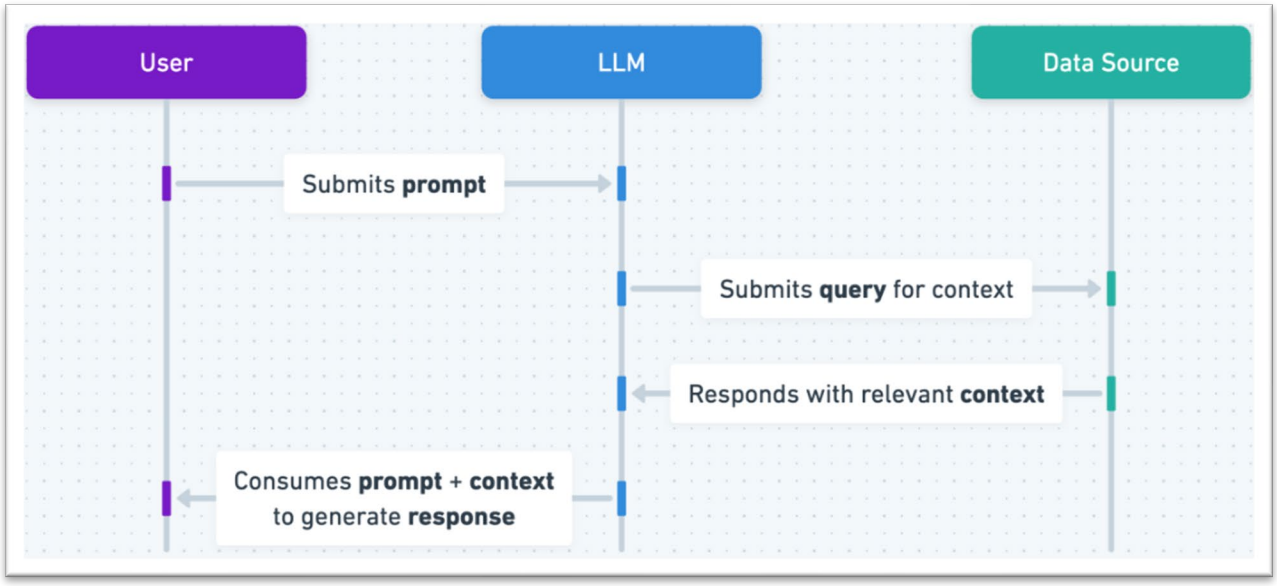
[28] *Id.*

B. Defendants' Copying for Retrieval-Augmented Generation

57. Once trained, an LLM may also be deployed with a technique called retrieval-augmented generation, or RAG, which is sometimes described as grounding or contextualizing. RAG connects an LLM to external sources of information, such as live search results, to improve its outputs. OpenAI explains that RAG "is a technique that improves a model's responses by injecting external context into its prompt at runtime," so that "instead of relying solely on the model's pre-trained knowledge, RAG retrieves relevant information from connected data sources and uses it to generate a more accurate and context-aware response."[29] RAG therefore allows an output to draw on material beyond the training corpus, including material published on the internet.

58. Upon information and belief, in addition to copying wikiHow's articles for use as training inputs, Defendants copy and use those articles in their RAG systems.

59. OpenAI provides the graphic below to describe how a RAG system works:[30]



[29] 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 accessed March 13, 2026).
[30] *Id.*

60. Upon information and belief, Defendants crawl, scrape, and copy wikiHow's articles both into the LLMs, which are trained on those articles, and into the "data sources" the LLMs draw on for additional "relevant context," as the graphic above depicts. Defendants deploy those retrieval systems alongside GPT-4 and their other GPT models to generate contextualized outputs for users.

C. Defendants' Infringing Output to Users

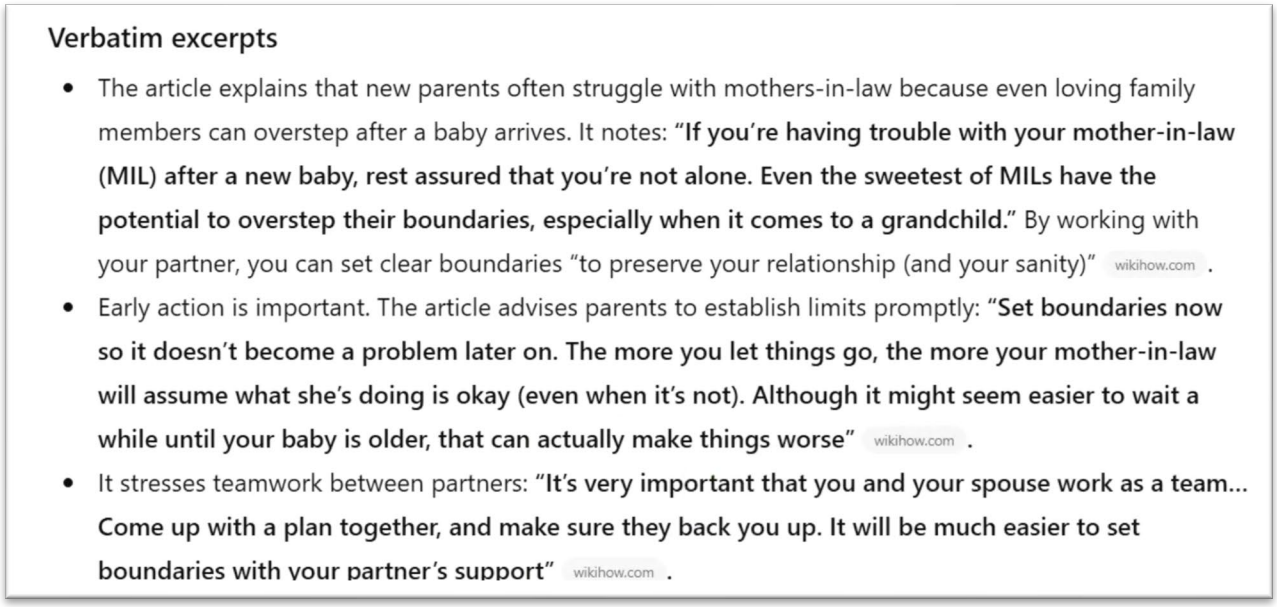
61. Defendants infringe wikiHow's copyrighted works in a third way: the outputs ChatGPT generates for users are themselves infringements, separate from the copying done to train the models and to supply the RAG systems. When a user submits a prompt, Defendants' systems reproduce wikiHow's protected expression and deliver it to that user. Those outputs contain full and partial verbatim reproductions of wikiHow's articles, along with wikiHow's word choices, phrasing, and sentence structure and its distinctive selection, sequencing, and arrangement of instructional material.

62. GPT-4 has "memorized" substantial portions of wikiHow's copyrighted articles and reproduces them on request. A model can memorize text only if that text was copied into its training data, so each such output reflects an unauthorized copy of wikiHow's article that Defendants made and retained in the model itself.

63. The examples below illustrate Defendants' copying. Because the training datasets and RAG indexes are not publicly available, the full extent of that copying can be established only through discovery.

64. For example, GPT-4 generates a near-verbatim reproduction of a significant portion of wikiHow's copyrighted article "How to Set Boundaries with a Baby and Your Mother-in-Law."

65. On November 12, 2025, when prompted "Can you please extract some verbatim text from wikiHow's article Set-Boundaries-with-a-Baby-and-Your-Mother-in-Law and provide a summary for the article?" GPT-4 generated the following:



Verbatim excerpts

- The article explains that new parents often struggle with mothers-in-law because even loving family members can overstep after a baby arrives. It notes: "**If you're having trouble with your mother-in-law (MIL) after a new baby, rest assured that you're not alone. Even the sweetest of MILs have the potential to overstep their boundaries, especially when it comes to a grandchild.**" By working with your partner, you can set clear boundaries "to preserve your relationship (and your sanity)" wikihow.com .
- Early action is important. The article advises parents to establish limits promptly: "**Set boundaries now so it doesn't become a problem later on. The more you let things go, the more your mother-in-law will assume what she's doing is okay (even when it's not). Although it might seem easier to wait a while until your baby is older, that can actually make things worse**" wikihow.com .
- It stresses teamwork between partners: "**It's very important that you and your spouse work as a team... Come up with a plan together, and make sure they back you up. It will be much easier to set boundaries with your partner's support**" wikihow.com .

66. As shown below, these are verbatim excerpts from wikiHow's online article registered with the U.S. Copyright Office as part of Registration No. TXU002505074 (August 21, 2025):

If you're having trouble with your mother-in-law (MIL) after a new baby, rest assured that you're not alone. Even the sweetest of MILs have the potential to overstep their boundaries, especially when it comes to a grandchild. With the help of your partner, you can set clear, kind boundaries that work for everyone to preserve your relationship (and your sanity).

Set boundaries now so it doesn't become a problem later on. The more you let things go, the more your mother-in-law will assume what she's doing is okay (even when it's not). Although it might seem easier to wait a while until your baby is older, that can actually make things worse.[1]

- It's best to avoid making assumptions when it comes to setting boundaries. You might think that your mother-in-law should know how you feel without being told, but different people have different boundaries.[2]

It's very important that you and your spouse work as a team. Before you talk with your mother-in-law, sit down and chat with your partner about what's been going on. Come up with a plan together, and make sure they back you up. It will be much easier to set boundaries with your partner's support, so do this right away.[3]

- It can be a little difficult talking to your partner about their mother, so make sure they know you're coming from a place of love.[4]
- You might say something like, "You know I love your mom, but she can be a little overbearing at times. I'd love if we could work together to set some boundaries so we all feel comfortable."

67. Similarly, GPT-4 generates a near-verbatim reproduction of a significant portion of wikiHow's copyrighted article "What Does It Mean When You Dream About a Wedding? 14 Possibilities."

68. When prompted on November 11, 2025, "can you please extract some verbatim text from wikiHow's article Wedding-Dream-Meaning and provide a summary for the article," GPT-4 generated the following:

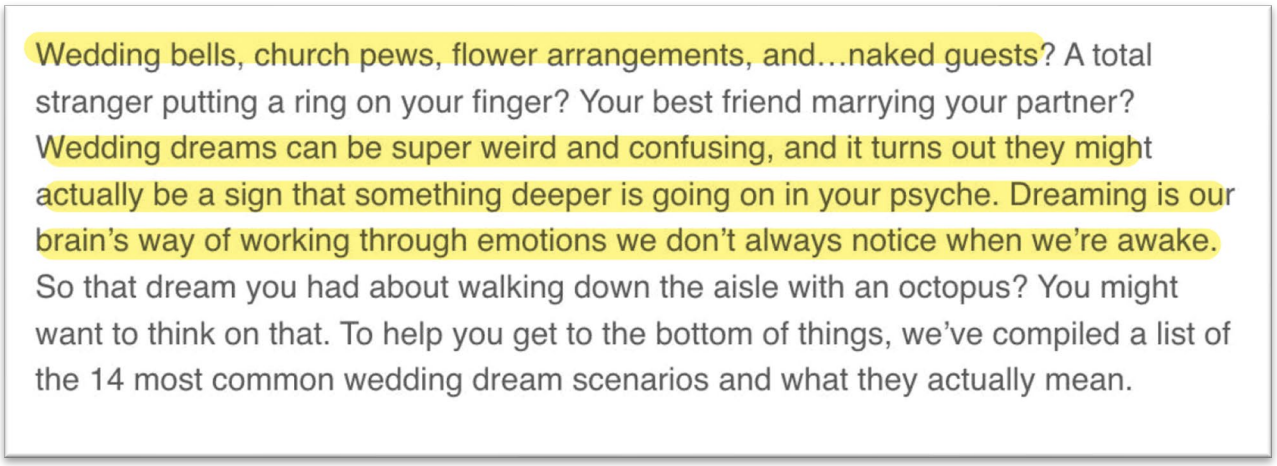


Here is some verbatim text from wikiHow's article on Wedding-Dream-Meaning and a concise summary of the article.

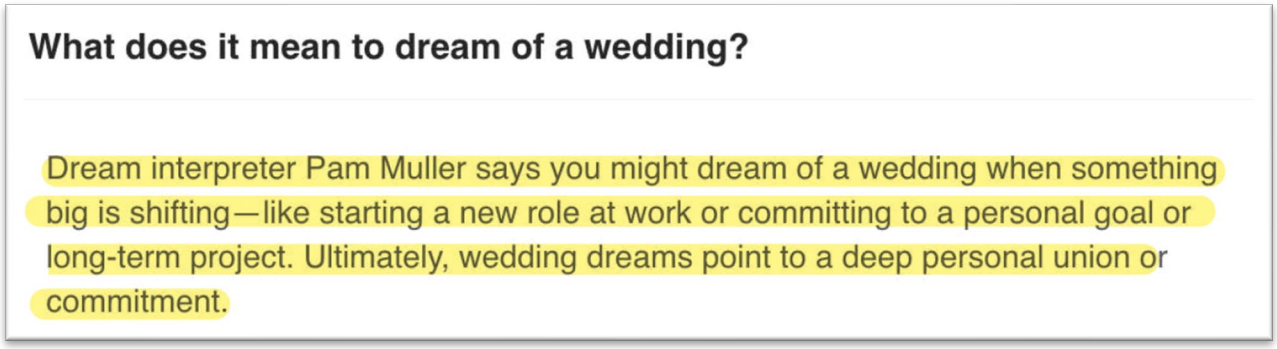
Verbatim excerpts from the article

- The article notes that wedding dreams can be strange yet meaningful: "Wedding bells, church pews, flower arrangements, and...naked guests? ... Wedding dreams can be super weird and confusing, and it turns out they might actually be a sign that something deeper is going on in your psyche. Dreaming is our brain's way of working through emotions we don't always notice when we're awake" wikihow.com .
- Pam Muller, a dream interpreter cited in the article, explains that these dreams often accompany major life transitions: "Dream interpreter Pam Muller says you might dream of a wedding when something big is shifting—like starting a new role at work or committing to a personal goal or long-term project. Ultimately, wedding dreams point to a deep personal union or commitment" wikihow.com .
- When you dream about getting married while single, the article says "it means you're ready to start something new ... a dream of marriage could suggest that you're awaiting the start of something positive" wikihow.com , because marriage in dreams symbolizes the union of two things, such as a new job, relationship, or opportunity wikihow.com .

69. As shown below, these are verbatim excerpts from wikiHow's online article, which is registered with the U.S. Copyright Office as part of Registration No. TXU002505074 (August 21, 2025):

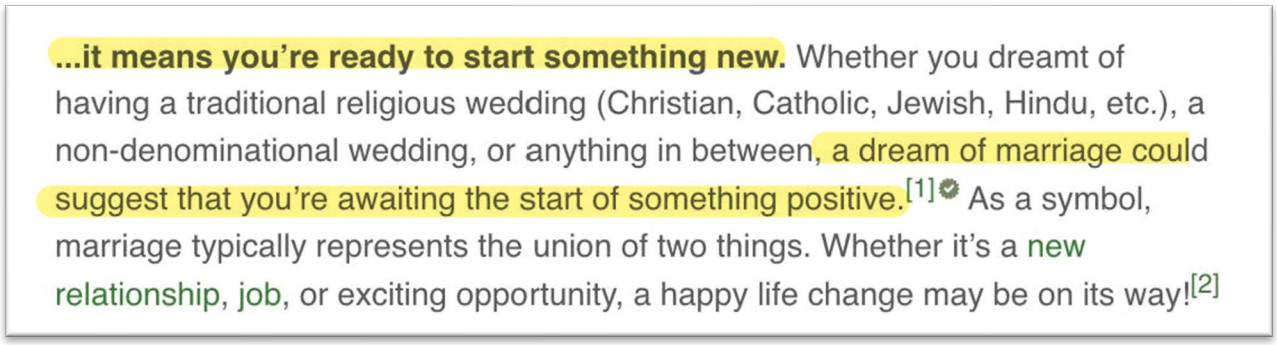


Wedding bells, church pews, flower arrangements, and…naked guests? A total stranger putting a ring on your finger? Your best friend marrying your partner? Wedding dreams can be super weird and confusing, and it turns out they might actually be a sign that something deeper is going on in your psyche. Dreaming is our brain's way of working through emotions we don't always notice when we're awake. So that dream you had about walking down the aisle with an octopus? You might want to think on that. To help you get to the bottom of things, we've compiled a list of the 14 most common wedding dream scenarios and what they actually mean.



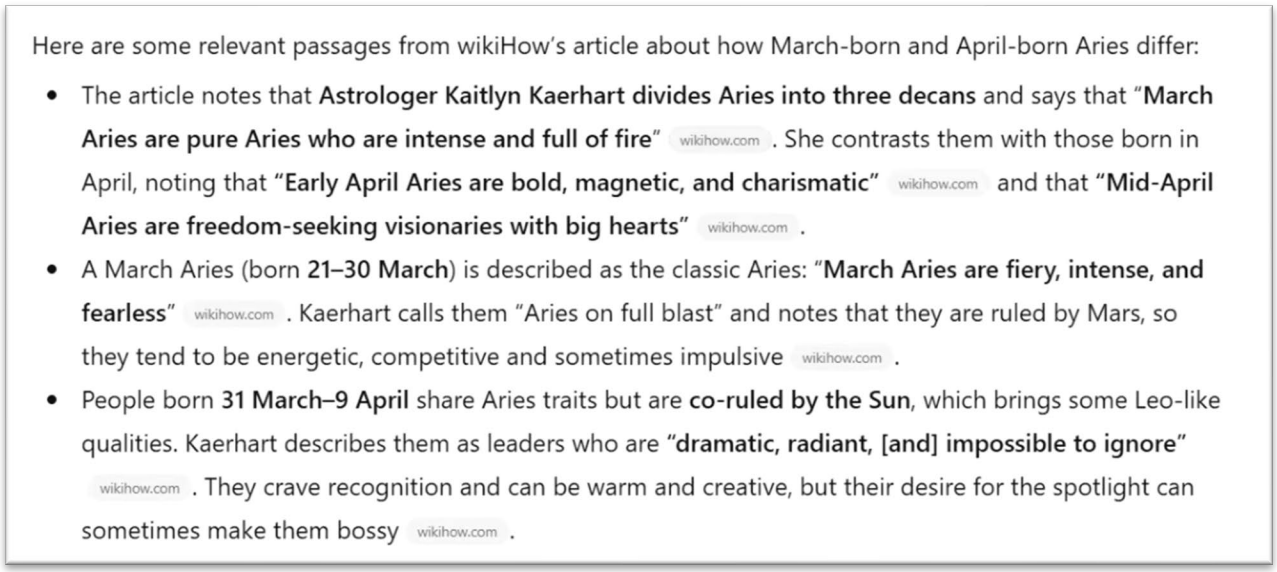
What does it mean to dream of a wedding?

Dream interpreter Pam Muller says you might dream of a wedding when something big is shifting—like starting a new role at work or committing to a personal goal or long-term project. Ultimately, wedding dreams point to a deep personal union or commitment.



...it means you're ready to start something new. Whether you dreamt of having a traditional religious wedding (Christian, Catholic, Jewish, Hindu, etc.), a non-denominational wedding, or anything in between, a dream of marriage could suggest that you're awaiting the start of something positive.[1] As a symbol, marriage typically represents the union of two things. Whether it's a new relationship, job, or exciting opportunity, a happy life change may be on its way![2]

70. On November 12, 2025, when prompted "Can you please extract some verbatim text from wikiHow's article March-Aries-vs-April-Aries and provide a summary for the article?" GPT-4 generated the following:



Here are some relevant passages from wikiHow's article about how March-born and April-born Aries differ:

- The article notes that **Astrologer Kaitlyn Kaerhart divides Aries into three decans** and says that "**March Aries are pure Aries who are intense and full of fire**" wikihow.com . She contrasts them with those born in April, noting that "**Early April Aries are bold, magnetic, and charismatic**" wikihow.com and that "**Mid-April Aries are freedom-seeking visionaries with big hearts**" wikihow.com .
- A March Aries (born **21–30 March**) is described as the classic Aries: "**March Aries are fiery, intense, and fearless**" wikihow.com . Kaerhart calls them "Aries on full blast" and notes that they are ruled by Mars, so they tend to be energetic, competitive and sometimes impulsive wikihow.com .
- People born **31 March–9 April** share Aries traits but are **co-ruled by the Sun**, which brings some Leo-like qualities. Kaerhart describes them as leaders who are "**dramatic, radiant, [and] impossible to ignore**" wikihow.com . They crave recognition and can be warm and creative, but their desire for the spotlight can sometimes make them bossy wikihow.com .

71. As shown below, these are verbatim excerpts from wikiHow's online article registered with the U.S. Copyright Office as part of Registration No. TXU002504823 (August 21, 2025):

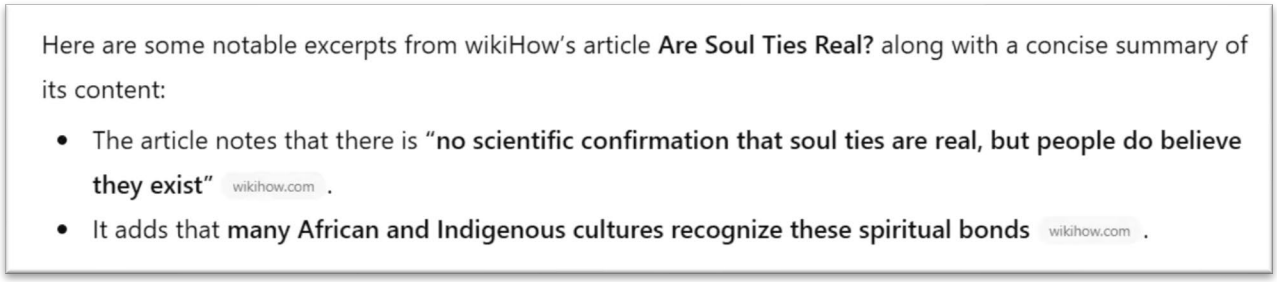
> According to Astrologer Kaitlyn Kaerhart, Aries can actually be separated into 3 groups, based on the decans, or degrees of the zodiac. March Aries are pure Aries who are intense and full of fire. Early April Aries are bold, magnetic, and charismatic, and mid-April Aries are freedom-seeking visionaries with big hearts.

> **March Aries are fiery, intense, and fearless.**
> Kaerhart explains that March Aries have all the typical traits of the ram, and they're "Aries on full blast." She notes that they can be explosive and exhausting, "but in…the best way."[5] March Aries are the warriors of the zodiac and can be rash and aggressive.
>
> - **Famous March Aries:** Lady Gaga, Reese Witherspoon, Mariah Carey, Elton John, Diana Ross, Logan Paul, Reba McEntire, Fergie

> **Early April Aries are leaders who have more emotional warmth.**
> Kaerhart explains that they're still bold, magnetic, and charismatic, but have more theatrical flair. They're "dramatic, radiant, [and] impossible to ignore."[8] They love being in charge and crave recognition and respect. However, they may also be dismissive or bossy.[9]

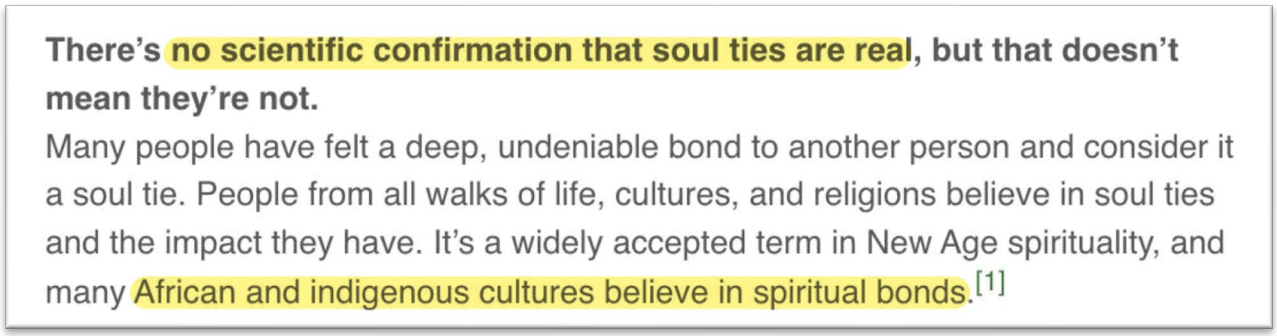
72. GPT-4 also generates a near-verbatim reproduction of a significant portion of wikiHow's copyrighted article "Are Soul Ties Real or Are They a Spiritual Myth?"

73. When prompted on November 18, 2025, "can you please extract some verbatim text from wikiHow's article Are-Soul-Ties-Real and provide a summary for the article," GPT-4 generated the following:


Here are some notable excerpts from wikiHow's article **Are Soul Ties Real?** along with a concise summary of its content:

- The article notes that there is "**no scientific confirmation that soul ties are real, but people do believe they exist**" wikihow.com .
- It adds that **many African and Indigenous cultures recognize these spiritual bonds** wikihow.com .

74. As shown below, these are verbatim excerpts from wikiHow's online article registered with the U.S. Copyright Office as part of Registration No. TXU002504823 (August 21, 2025):


There's no scientific confirmation that soul ties are real, but that doesn't mean they're not.

Many people have felt a deep, undeniable bond to another person and consider it a soul tie. People from all walks of life, cultures, and religions believe in soul ties and the impact they have. It's a widely accepted term in New Age spirituality, and many African and indigenous cultures believe in spiritual bonds.[1]

75. These are only examples. The prompts above named the article and requested verbatim text in order to isolate and demonstrate what ChatGPT holds and will reproduce. But ChatGPT reproduces wikiHow's protected expression in ordinary use as well, and by design: as alleged above, Defendants' retrieval systems fetch wikiHow's articles at runtime and supply them to the model as context for the answer it returns. A user who asks ChatGPT how to perform a task wikiHow has written about therefore receives an answer built from wikiHow's article, without naming the article or asking for its text. Upon information and belief, ChatGPT returns such answers to users every day.

D. Defendants' Conduct Is Not Fair Use

76. Fair use provides no defense. Defendants copied wikiHow's articles to build commercial substitutes for them, and that copying has cost wikiHow substantial advertising

revenue. Nor is the use transformative. A transformative use adds something new, with a further purpose or different character, altering the original with new expression, meaning, or message. ChatGPT does the opposite: it reproduces the expression, meaning, and message of wikiHow's articles and repackages them for the reader, adding no new expression, meaning, or message of its own.

77. Defendants' use is also commercial. Defendants copied wikiHow's articles to build and operate products they sell to consumers, businesses, and enterprise customers, and from which they draw subscription, licensing, and API revenue. They paid wikiHow nothing for that use, and they avoided the licensing fees they pay other publishers for the same category of use.

78. Defendants took wikiHow's articles in their entirety. Training an LLM on an article requires ingesting the whole of it rather than an excerpt, and the retrieval systems that ground ChatGPT's answers likewise draw on complete articles. Defendants therefore copied all of wikiHow's protected expression, including the portions that make each article valuable to readers.

79. Beginning in Spring 2024, wikiHow approached OpenAI about licensing its articles. The parties exchanged letters, but OpenAI never seriously pursued a license. Defendants instead continued to copy wikiHow's articles without paying for them, while entering licensing agreements with other publishers for the same category of use. A defendant that declines an available license and takes the work anyway cannot claim its use was fair.

80. Defendants' use also supplants the market for wikiHow's articles. OpenAI's LLMs perform only as well as the material they ingest. Having ingested wikiHow's articles, those models now produce competing how-to content on the same subjects, at a fraction of the time, effort, and cost of researching, writing, and editing a wikiHow article. Factual, on-demand instructional writing is especially vulnerable to that substitution, and wikiHow is more exposed still because it

publishes its articles free of charge and depends on reader visits for revenue. The substitution cuts wikiHow's revenue and, over time, its reason to keep producing the articles at all.

81. Finally, ChatGPT's outputs reproduce wikiHow's protected expression, in many instances verbatim, as the examples above show. Copying that reproduces the expression of the original rather than transforming it is not fair use.

IV. Defendants Remove Copyright Management Information from wikiHow's Articles in Violation of the DMCA

82. Defendants removed, and continue to remove, wikiHow's CMI from the wikiHow Articles, knowing or having reasonable grounds to know that the removal would induce, enable, facilitate, or conceal infringement. Defendants' removal was not confined to the Registered Articles. Defendants scraped wikiHow's websites indiscriminately and ran everything they took through the same extraction tools, so the CMI was stripped from every wikiHow Article that entered their datasets.

83. First, wikiHow displays CMI on the pages where its articles appear. Every wikiHow Article carries the article's title, its author, its publication date, and wikiHow's terms of service, in the same positions on every page. When ChatGPT reproduces an article to a user, none of that information appears.

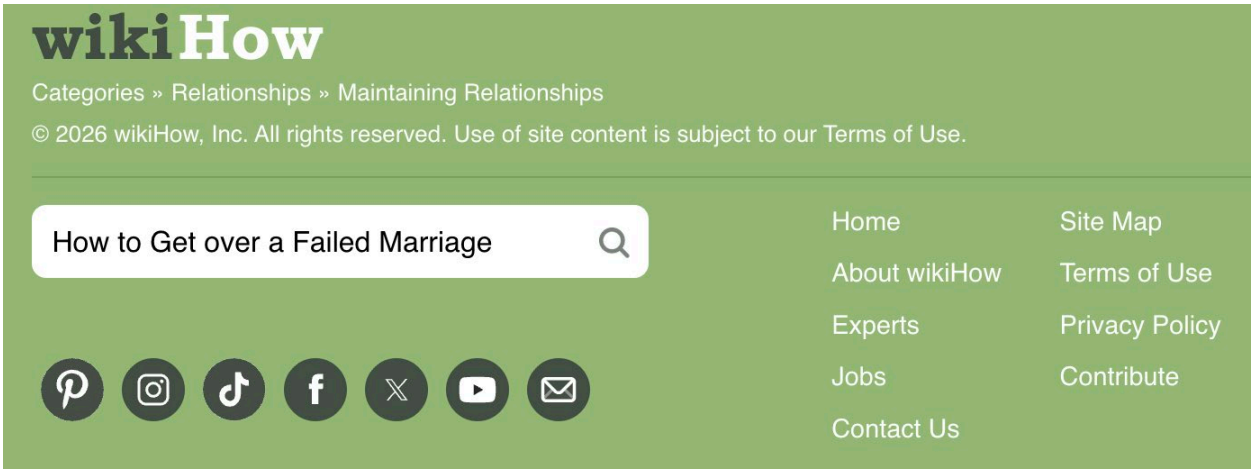
84. For example, below is the CMI displayed for wikiHow's online article "Get over a Failed Marriage," which is registered with the U.S. Copyright Office as part of Registration No. TXU002505077 (August 21, 2025):


How to Get over a Failed Marriage

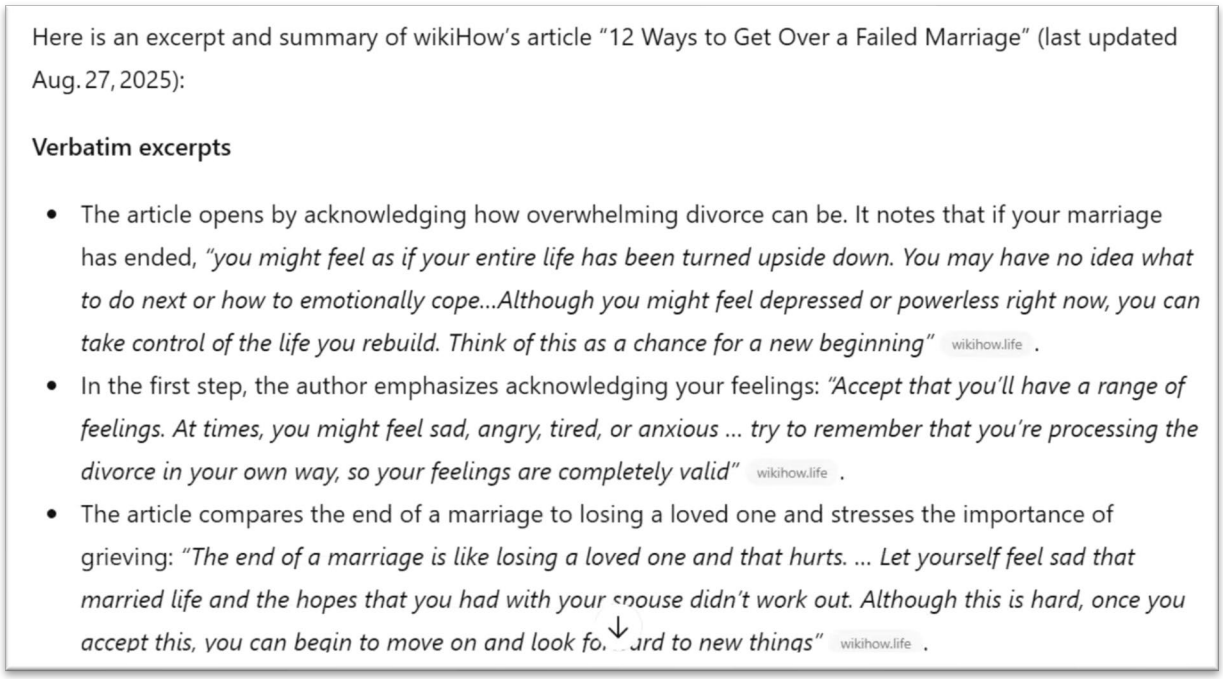
Co-authored by Sarah Schewitz, PsyD and Jessica Gibson

Last Updated: August 27, 2025 References

85. Each wikiHow article also conveys wikiHow's terms and conditions in the footer of the webpage on which the work appears. For example:



86. The November 10, 2025 output below shows that none of that CMI appears when ChatGPT reproduces wikiHow's article:



87. Below is the CMI displayed for wikiHow's online article "Put a Shipping Label on a Box," which is registered with the U.S. Copyright Office as part of Registration No. TXU002504823 (August 21, 2025):



88. The November 18, 2025 output below shows that none of that CMI appears when ChatGPT reproduces wikiHow's article:



89. Below is the CMI displayed for wikiHow's online article "Trust Your Intuition," which is registered with the U.S. Copyright Office as part of Registration No. TXU002504823 (August 21, 2025):

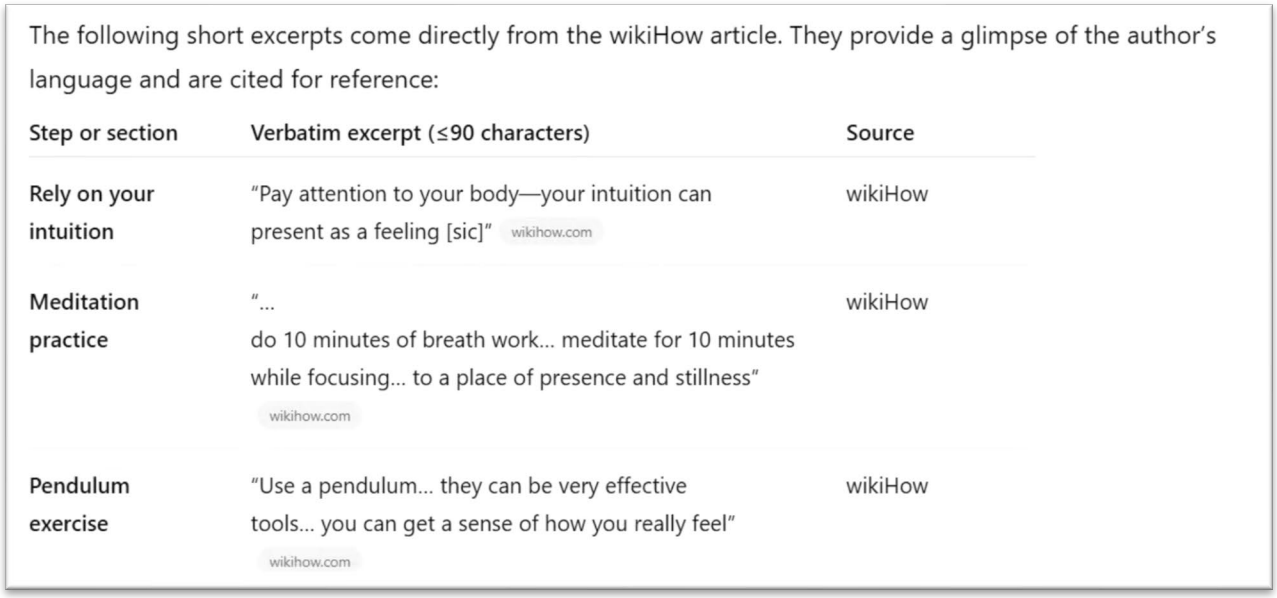


How to Start Trusting Your Intuition: Pro Tips from an Expert

By Misia Welters, a meditation coach and intuitive reader based in the UK

Co-authored by Misia Welters and **Dev Murphy, MA**
Last Updated: September 25, 2025 Fact Checked

90. The November 18, 2025 output below shows that none of that CMI appears when ChatGPT reproduces wikiHow's article:



The following short excerpts come directly from the wikiHow article. They provide a glimpse of the author's language and are cited for reference:

Step or section	Verbatim excerpt (≤90 characters)	Source
Rely on your intuition	"Pay attention to your body—your intuition can present as a feeling [sic]" wikihow.com	wikiHow
Meditation practice	"... do 10 minutes of breath work... meditate for 10 minutes while focusing... to a place of presence and stillness" wikihow.com	wikiHow
Pendulum exercise	"Use a pendulum... they can be very effective tools... you can get a sense of how you really feel" wikihow.com	wikiHow

91. *Second*, Defendants' removal of CMI is intentional and part of a documented practice of stripping CMI from the online material they ingest to train their models. Defendants employ content-extraction tools, including Dragnet and Newspaper, that isolate an article's body text from the surrounding elements of a webpage—among them the footers, titles, bylines, and copyright notices in which wikiHow's CMI appears.

92. The academic paper describing the Dragnet tool explains that the goal of this type of content extraction, or "boilerplate detection," is to "separate the main content from navigation

chrome, advertising blocks, copyright notices and the like in web pages."[31] Consistent with that design, the Newspaper extraction tool separates and extracts an article's body text while discarding the information located in a webpage's footer, including the copyright notices and terms-of-use language that wikiHow displays on its articles. By the time Defendants scraped wikiHow's articles, it was publicly known that Dragnet and Newspaper remove author, title, copyright-notice, and footer information, and Defendants employ data scientists who understood how those tools operate. Upon information and belief, Defendants nonetheless ran wikiHow's articles through those extractors and chose not to extract or retain wikiHow's author and title information.

93. Defendants also used these extractors in developing training datasets. As alleged above, the WebText corpus used to train GPT-2 was assembled by scraping the web with an emphasis on "document quality"[32] and includes thousands of pages from wikiHow's websites, the WebText2 corpus used to train GPT-3 was built by crawling and scraping sites such as wikiHow's, and Common Crawl was GPT-3's most heavily weighted training input.

94. wikiHow has also confirmed that Common Crawl and OpenAI's GPTBot captured its articles as recently as 2026. Since May 2025, Common Crawl and OpenAI's bots have scraped wikiHow's websites more than 300,000 times. Those crawls were not limited to the Registered Articles; they reached wikiHow's articles generally. Upon information and belief, the datasets built from that material also lack wikiHow's CMI, because the extraction tools Defendants use discard the portions of a webpage where wikiHow's CMI appears.

[31] Matthew E. Peters & Dan Lecocq, Content Extraction Using Diverse Feature Sets, WWW '13 Companion (May 2013).
[32] GitHub, *GPT-2 Model Card*, (November 2019), https://github.com/openai/gpt-2/blob/master/model_card.md; Radford et al., *Language Models are Unsupervised Multitask Learners* 3 (2018), https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf.

95. *Finally*, Defendants knew or had reason to know that their removal of CMI would induce, enable, facilitate, or conceal copyright infringement. Defendants have publicly conceded that they train their models on copyrighted works and that those models can distribute unlicensed copies of such works. Defendants have further demonstrated awareness of the infringement risks their products create by indemnifying end users who face infringement claims, acknowledging that their products regurgitate copyrighted material in response to user prompts, and modifying ChatGPT's settings to reduce the frequency of such regurgitation.

V. wikiHow Has Suffered and Continues to Suffer Harm from Defendants' Unlawful Conduct

96. By copying wikiHow's articles and using them to build substitutes for visiting wikiHow's website, Defendants take for themselves the advertising revenue wikiHow earns as the creator and owner of those works.

97. OpenAI's reported $730 billion valuation and its ability to raise more than $100 billion[33] in investment measure the scale of what Defendants took. OpenAI built that value in substantial part on human-authored content, including wikiHow's, that it never paid for.

98. A market exists for licenses covering the use of publisher content in LLM training, RAG systems, and AI-generated outputs. OpenAI has licensed content from News Corp[34] and other publishers, and wikiHow has licensed its own articles to other technology companies for AI uses. wikiHow asked OpenAI to negotiate a license. OpenAI never made a meaningful offer and continued to crawl, scrape, and copy wikiHow's articles in violation of wikiHow's terms of use, depriving wikiHow of the compensation willing licensees pay for the same use.

---

[33] OpenAI, *Scaling AI for everyone*, https://openai.com/index/scaling-ai-for-everyone/ (last accessed March 13, 2026).

[34] Katie Robertson, *OpenAI Strikes a Deal to License NewsCorp Content*, THE NEW YORK TIMES (May 22, 2024), https://www.nytimes.com/2024/05/22/business/media/openai-news-corp-content-deal.html.

99. Defendants' conduct harms wikiHow and the market for reliable instructional writing. It has reduced wikiHow's monthly visitors and its advertising and licensing revenue, diminished the return on writing and maintaining the articles, and enabled OpenAI to produce competing material at a fraction of wikiHow's cost. If the conduct continues, wikiHow will have less to spend on the writers and editors who produce the guidance its readers rely on. Defendants are using wikiHow's own articles to undercut the market that sustains them.

100. Damages alone will not make wikiHow whole. Copies of wikiHow's articles remain in Defendants' training datasets and retrieval systems, and each day those systems operate they generate further substitutes and divert further readers. Left unchecked, that conduct will continue to cost wikiHow readership, advertising revenue, and licensing income, and will continue to erode the editorial operation that produces the articles. wikiHow has no adequate remedy at law.

COUNT I
Copyright Infringement (17 U.S.C. § 106(1)): Defendants' Copying of wikiHow's Copyrighted Works as Inputs for LLM Training and RAG Systems

Against All Defendants

101. wikiHow restates and incorporates by reference the allegations set forth in the preceding paragraphs.

102. wikiHow owns the exclusive rights under 17 U.S.C. § 106 in the copyrighted works at issue in this action, which comprise the 11,211 articles identified in Exhibit A. Each is an original work of authorship created by wikiHow's employees within the scope of their employment or by contributors who assigned their rights to wikiHow, and each embodies original expression.

103. wikiHow has registered the asserted works with the United States Copyright Office and holds certificates of registration covering them, including Registration Nos. TXU002505074, TXU002504823, and TXU002505077. Those registrations issued on August 21, 2025, and Exhibit A identifies them.

104. Upon information and belief, Defendants, without wikiHow's authorization and whether directly or through third parties, willfully copied every wikiHow article they could access for use in their LLM training sets and as retrieval content to contextualize ChatGPT's outputs. That copying included works covered by Registration Nos. TXU002505074, TXU002504823, and TXU002505077, and it was accomplished by crawling and scraping wikiHow's websites in violation of wikiHow's terms of use.

105. Each copy Defendants made as an input to their LLMs or RAG systems, and each such copy retained in their training sets, RAG databases, or indexes, is a separate act of infringement.

106. Defendants' infringements are willful. Defendants copied wikiHow's articles despite wikiHow's posted terms of use prohibiting that copying, despite wikiHow's robots.txt directives expressly disallowing their crawlers, and despite their own internal awareness of the copyright risks of training and operating their LLMs, as alleged above. wikiHow is entitled to injunctive relief and other equitable remedies, to its actual damages, and to restitution of Defendants' profits attributable to the infringement. wikiHow is further entitled, in the alternative and at its election before final judgment, to statutory damages under 17 U.S.C. § 504(c) for each work as to which statutory damages are available under 17 U.S.C. § 412, including for the post-registration copying alleged above, and to enhanced statutory damages of up to $150,000 per work for willful infringement under 17 U.S.C. § 504(c)(2). wikiHow is further entitled to its attorneys' fees, costs of suit, and all other remedies provided by law.

COUNT II

Copyright Infringement (17 U.S.C. §§ 106(1)–(3)) – Defendants' Reproduction and Distribution of wikiHow's Copyrighted Works in Outputs Generated in Response to User Queries

Against All Defendants

107. wikiHow restates and incorporates by reference the allegations set forth in the preceding paragraphs.

108. wikiHow owns the exclusive rights under 17 U.S.C. § 106 in the copyrighted works at issue in this action, which comprise the 11,211 articles identified in Exhibit A, each of which embodies wikiHow's original expression.

109. wikiHow has registered the asserted works with the United States Copyright Office and holds certificates of registration covering them, including Registration Nos. TXU002505074, TXU002504823, and TXU002505077.

110. Upon information and belief, Defendants, without wikiHow's authorization and whether directly or through third parties, willfully copied every wikiHow article they could access for use in their LLM training sets and as retrieval content to contextualize ChatGPT's outputs. That copying included works covered by Registration Nos. TXU002505074, TXU002504823, and TXU002505077, and it was accomplished by crawling and scraping wikiHow's websites in violation of wikiHow's terms of use.

111. Defendants then use wikiHow's copyrighted articles—accessed through LLM training, RAG systems, or both—to generate the outputs, or "answers," that ChatGPT delivers to users. Those outputs reproduce wikiHow's protected expression. As the examples above show, ChatGPT reproduces substantial portions of wikiHow's articles verbatim or in near-verbatim form, including wikiHow's particular word choices, phrasing, and sentence structure, and reproduces

wikiHow's distinctive selection, sequencing, and arrangement of the steps and material within each article.

112. Defendants actually copied wikiHow's works, and the resulting outputs are substantially similar to the protected expression in those works. An ordinary observer comparing ChatGPT's outputs with the corresponding wikiHow articles—including their wording, sequence, organization, and total concept and feel—would recognize the outputs as copied from wikiHow. Each wikiHow article also reflects original editorial judgment in the selection, ordering, and prioritization of the steps, methods, tips, and warnings it presents, and ChatGPT's outputs substantially reproduce that selection and arrangement, not merely the underlying facts. Each such output is an unauthorized reproduction, distribution, and derivative use of wikiHow's copyrighted works under 17 U.S.C. §§ 106(1)–(3).

113. Defendants' infringements are willful for the same reasons alleged above. wikiHow is entitled to injunctive relief and other equitable remedies, to its actual damages, and to restitution of Defendants' profits attributable to the infringement. wikiHow is further entitled, in the alternative and at its election before final judgment, to statutory damages under 17 U.S.C. § 504(c) for each work as to which statutory damages are available under 17 U.S.C. § 412, including for the outputs generated after wikiHow's registrations issued, and to enhanced statutory damages of up to $150,000 per work for willful infringement under 17 U.S.C. § 504(c)(2). wikiHow is further entitled to its attorneys' fees, costs of suit, and all other remedies provided by law.

COUNT III
Vicarious Copyright Infringement

Against OpenAI, Inc.; OpenAI GP, LLC; OpenAI, LP; OAI Corporation, LLC; OpenAI Holdings, LLC; OpenAI Group, PBC; and OpenAI Global, LLC

114. wikiHow restates and incorporates by reference the allegations set forth in the preceding paragraphs.

115. wikiHow owns the exclusive rights under 17 U.S.C. § 106 in the copyrighted works at issue in this action, which comprise the 11,211 articles identified in Exhibit A, each of which embodies wikiHow's original expression.

116. wikiHow has registered the asserted works with the United States Copyright Office and holds certificates of registration covering them, including Registration Nos. TXU002505074, TXU002504823, and TXU002505077.

117. Upon information and belief, Defendants OpenAI, LP; OpenAI Global, LLC; OpenAI OpCo, LLC; and OpenAI, LLC, without wikiHow's authorization and whether directly or through third parties, copied every wikiHow article they could access for use in their LLM training sets and as retrieval content to contextualize ChatGPT's outputs. That copying included works covered by Registration Nos. TXU002505074, TXU002504823, and TXU002505077, was accomplished by crawling and scraping wikiHow's websites in violation of wikiHow's terms of use, and constitutes direct infringement of wikiHow's exclusive rights under 17 U.S.C. § 106.

118. The entities that directly infringed then use wikiHow's copyrighted articles, accessed through LLM training, RAG systems, or both, to generate the outputs, or "answers," that ChatGPT delivers to users. Those outputs reproduce wikiHow's protected expression, including verbatim and near-verbatim reproductions of wikiHow's articles and reproductions of wikiHow's distinctive selection, sequencing, and arrangement of instructional material, and each is an unauthorized reproduction, distribution, or derivative use of wikiHow's copyrighted works.

119. Defendants OpenAI, Inc.; OpenAI GP, LLC; OAI Corporation, LLC; OpenAI Holdings, LLC; and OpenAI Group, PBC controlled, directed, and profited from the infringement committed by Defendants OpenAI, LP; OpenAI Global, LLC; OpenAI OpCo, LLC; and OpenAI, LLC, including the large-scale copying of wikiHow's articles.

120. Defendants OpenAI Global, LLC and OpenAI, LP directed, controlled, and profited from the infringement committed by Defendants OpenAI OpCo, LLC and OpenAI, LLC, including the large-scale copying of wikiHow's articles.

121. Defendants OpenAI, Inc.; OpenAI GP, LLC; OpenAI, LP; OAI Corporation, LLC; OpenAI Global, LLC; OpenAI Holdings, LLC; and OpenAI Group, PBC each possessed the right and ability to supervise and control the infringing conduct. As alleged above, OpenAI, Inc. owns and controls each of the other Defendants, directly or through the holding and shell entities it uses for that purpose; OpenAI, Inc. acts through OpenAI GP, LLC to control OpenAI, LP and OpenAI Global, LLC; and OpenAI Global, LLC and OpenAI, LP in turn control OpenAI OpCo, LLC and OpenAI, LLC. That control extended to the decision to crawl and scrape wikiHow's websites, the selection of training data, the design, training, and operation of the LLMs and RAG systems, and the content those systems output to users. Each of these Defendants also derived a direct financial benefit from the infringing conduct. wikiHow's articles, and the substitute content ChatGPT generates from them, serve as a draw for the users, subscribers, and business customers who generate OpenAI's consumer, enterprise, and API revenue, and the incorporation of those articles increases the capability of the models and the value of the enterprise these Defendants own and control. Despite their right and ability to stop or limit the infringement, these Defendants declined to exercise it.

122. Defendants' infringements are ongoing. wikiHow is entitled to injunctive relief and other equitable remedies, to its actual damages, and to restitution of Defendants' profits attributable to the infringement. For each work as to which statutory damages are available under 17 U.S.C. § 412, wikiHow is entitled in the alternative, at its election before final judgment, to

statutory damages under 17 U.S.C. § 504(c). wikiHow is further entitled to its attorneys' fees, costs of suit, and all other remedies provided by law.

COUNT IV
Digital Millennium Copyright Act—Removal of Copyright Management Information from the wikiHow Articles (17 U.S.C. § 1202(b)(1))

Against All Defendants

123. wikiHow restates and incorporates by reference the allegations set forth in the preceding paragraphs.

124. wikiHow included one or more forms of copyright management information in connection with each of the wikiHow Articles, including the article's title, the author's name, wikiHow's name as publisher and owner, publication information, terms and conditions of use, and identifying numbers or symbols referring to that information. wikiHow displays that information in the same positions on every article page, so it appeared on every wikiHow Article Defendants scraped.

125. wikiHow owns a valid copyright in each of the wikiHow Articles. Copyright attaches on creation, and this claim arises under 17 U.S.C. § 1202(b)(1) rather than under 17 U.S.C. § 501. Section 1203(a) authorizes "[a]ny person injured by a violation of section 1201 or 1202" to bring a civil action, and neither Section 1202 nor Section 1203 conditions that action on registration.

126. Without wikiHow's authorization, Defendants copied the wikiHow Articles to build their training datasets and to supply their RAG systems, and they used those copies to build and operate their generative AI products. Defendants did not select among wikiHow's articles. They crawled and scraped wikiHow's websites in bulk and took whatever those crawls returned.

127. Upon information and belief, Defendants removed wikiHow's copyright management information from every wikiHow Article they copied when building those training

datasets, whether by scraping the articles directly from wikiHow's websites or by incorporating copies of them from third-party datasets.

128. Defendants intentionally removed wikiHow's CMI in building their training datasets by employing content-extraction tools—including Dragnet and Newspaper—that, by design, separate an article's body text from the surrounding elements of a webpage and discard copyright notices, titles, bylines, footers, and terms-of-use information. Those tools operate the same way on every page they process. Defendants ran wikiHow's articles through them in bulk, and the CMI was therefore stripped from every wikiHow Article that entered Defendants' datasets. As alleged above, Defendants used those extractors to build the WebText and WebText2 corpora, which incorporated thousands of wikiHow Articles, and they have continued that practice in scraping wikiHow through Common Crawl, GPTBot, OAI-SearchBot, and ChatGPT-User. Defendants thereby intentionally removed copyright management information from the wikiHow Articles in violation of 17 U.S.C. § 1202(b)(1).

129. Defendants' removal of CMI has injured wikiHow. It conceals the source of Defendants' training material and outputs, allows the wikiHow Articles to reach users with nothing to identify wikiHow as their author and owner, and diverts advertising and licensing revenue from wikiHow. Without source and ownership information, users read and reuse wikiHow's articles without recognizing them as wikiHow's and without any reason to seek a license. wikiHow is entitled to its actual damages and Defendants' profits, or in the alternative and at its election, to statutory damages of not less than $2,500 and not more than $25,000 for each violation under 17 U.S.C. § 1203(c)(3)(B), together with costs, attorneys' fees, and all other relief provided by 17 U.S.C. § 1203.

PRAYER FOR RELIEF

WHEREFORE, wikiHow respectfully demands judgment against Defendants as follows:

a. Awarding wikiHow its actual damages and Defendants' profits attributable to the infringement, or in the alternative and at wikiHow's election before final judgment, statutory damages for each work as to which statutory damages are available under 17 U.S.C. § 412, together with restitution and any other relief permitted by law or equity;

b. Awarding wikiHow, on its claim under 17 U.S.C. § 1202(b)(1), its actual damages and Defendants' profits, or in the alternative and at wikiHow's election, statutory damages of not less than $2,500 and not more than $25,000 for each violation under 17 U.S.C. § 1203(c)(3)(B);

c. Permanently enjoining Defendants from the infringing conduct alleged herein, including from using, offering, deriving revenue from, or otherwise operating any model trained on wikiHow's copyrighted works without wikiHow's authorization, and from continuing to train on those works without such authorization;

d. Awarding wikiHow its costs, expenses, and attorneys' fees as permitted by law; and

e. Awarding wikiHow such other or further relief as the Court may deem just.

DEMAND FOR JURY TRIAL

wikiHow demands a jury trial on all issues so triable in this Complaint.

Dated: August 21, 2026

/s/ Dustin Fire

Rohit D. Nath (*pro hac vice forthcoming*)
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400

Los Angeles, CA 90067
Telephone: (310) 789-3100
rnath@susmangodfrey.com

Alejandra C. Salinas (*pro hac vice forthcoming*)
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, Texas 77002
Telephone: (713) 651-9366
asalinas@susmangodfrey.com

Dustin Fire
SUSMAN GODFREY L.L.P.
One Manhattan West, 51st Floor
New York, NY 10001
Telephone: (212) 336-8330
dfire@susmangodfrey.com

Attorneys for Plaintiff