Collaborative Student Work
AI COPYRIGHT LITIGATION V. LICENSING: AN EXPLORATION

by Lily Argyle, Tess Bradley, James Dearing, Hallie Feinman, Layla Ghiai, Gerado Gomez, Marie Gosserand, Holland Haney, Jamika Mack, Maddy Marlowe, Whitney McBay, Haley Pettingill, Isa Reyes-Klein, Nathan Schwartz, Jon Evan Smelley, Natalia Walcott and Tynesea Watts

This report seeks to understand the current state of litigation and licensing related to copyright cases related to artificial intelligence. Conducted as a classroom exercise at Tulane Law School, we thought others might find it useful as well. In short, there's a lot going on, and keeping up can be challenging.

INTRODUCTION

On November 30, 2022, we were introduced to ChatGPT.1 By 2023, we had not only OpenAI, but many other companies producing AI tools; with this major industry boom came a slew of lawsuits.2 However, what was unconventional was the slew of licensing deals between AI companies and media outlets/companies.3 Then, on February 11, 2025, we got the first summary judgment related to copyright infringement and unauthorized use of materials; in this case, a non-generative system using Thomson Reuters' Westlaw headnotes without authorization in a competing product.4 The world is moving both at a slow pace (litigation) and fast (licensing). And it's hard to keep up and digest the many moving parts.

This summary reviews the state of AI cases and licensing as of February 2025.5 Part I asks the question of how many AI copyright cases are there and explains why there is not quite a straightforward answer. Then, Part II looks at who are some of the defendants in the cases. We see certain companies being sued by multiple plaintiffs, and we are trying to get a sense of what's going on. Part III then turns to the cases themselves. We will try to give you a summary of

---

[1] *Introducing ChatGPT*, OPENAI (Nov. 30, 2022), https://openai.com/index/chatgpt/.
[2] Tremblay v. OpenAI, Inc., 716 F. Supp. 3d 772 (N.D. Cal. 2024).
[3] *See, e.g., Authors Guild Partners with Created by Humans to Empower Authors in the AI Era*, THE AUTHORS GUILD: PRESS RELEASES (Oct. 9, 2024), https://authorsguild.org/news/ag-partners-with-created-by-humans-to-empower-authors-in-ai-era/ (Working to streamline licensing, Created by Humans is a "pioneering platform that enables authors to license their work to AI developers" This new partnership attempts to "help protect and promote authors' rights.").
[4] Thomson Reuters v. ROSS, 694 F. Supp. 3d 467 (D. Del. 2023).
[5] This review was conducted by the students in the Advanced Copyright and Trademark Course at Tulane Law School, under the guidance of Dr. Elizabeth Townsend Gard. Each student is credited with the case/section they worked on. We created this collaborative document as a learning experience, and wanted to share what we found. All errors or omissions are the responsibility of the law students and Dr. Townsend Gard, and we apologize for errors and omissions. Please contact Dr. Townsend Gard at townsend@tulane.edu for any necessary revisions or edits.

some of the cases so far. Part IV then looks at a sample of secondary sources –
news stories and law review articles and gives a sample of how people are
approaching the question. Part V looks at the recent summary judgement in
Thomson v. ROSS, the first AI copyright case filed, and decided on February 11,
2025. Although not a generative AI case, could the approach of this district court
give us insight or help other cases with their strategies going forward? Part VI
looks at the rise of licensing deals in 2024. Part VII does a quick summing up and
speculation as of February 2025.

I.   HOW MANY AI COPYRIGHT CASES ARE THERE?[6]

As we begin this piece, we see different numbers reported. There's no
definitive list. In part, due to the high number of repeat defendants, these AI
copyright cases keep getting consolidated. For example, in December 2024,
OpenAI sought consolidation of eight ongoing copyright lawsuits against them.[7]
Taking that into account, the numbers do vary quite a bit.

In this section, we are providing resources for you to track the cases, and we
have found that viewing a number of sources proves helpful in the numbers
anomaly. A number of resources are tracking the cases as they work their way
through the courts. Some resources only concentrate on a handful of cases,
focusing on what their potential effect on copyright might be.[5] Other resources
are simply tracking AI cases, regardless of whether they are consolidated or not.
Some are doing more sophisticated summaries.

Let's take the month of December 2024. Copyright Alliance reports well
over 30 cases have been filed.[8] Wired reports 27 AI copyright cases. [9] The law
firm, Ropes and Gray, reports 15 *notable* cases.[10] On the other hand, Tech Policy
reports only 13 *notable* cases. [11] Let's look at and compare how some resources

---

[6] Authored by Tess Bradley, Tulane Law School, J.D. expected May 2025.

[7] Annelise Levy, *OpenAI Seeks to Centralize Copyright Lawsuits Against It*, BLOOMBERG
LAW (Dec. 5, 2024), https://news.bloomberglaw.com/ip-law/openai-to-seek-to-centralize-
eight-copyright-lawsuits-against-it.

[8] Keith Kupferschmeid, *Insights from Court Orders in AI Copyright Infringement Cases*,
COPYRIGHT ALLIANCE (Dec. 12, 2024), https://copyrightalliance.org/ai-copyright-
infringement-cases-insights/#1note.

[9] Kate Knibbs, *Every AI Copyright Lawsuit in the U.S., Visualized*, WIRED: BUSINESS (Dec.
19, 2024), https://www.wired.com/story/ai-copyright-case-tracker/.

[10] David M. McIntosh et al., *An End-of-the-Year Update to the Current State of AI
Copyright Litigation*, ROPES & GRAY (Dec. 17, 2024),
https://www.ropesgray.com/en/insights/alerts/2024/12/an-end-of-year-update-to-the-
current-state-of-ai-related-copyright-litigation.

[11] Bruce Barcott, *AI Lawsuits Worth Watching: A Curated Guide*, TECH POLICY PRESS (July
1, 2024), https://www.techpolicy.press/ai-lawsuits-worth-watching-a-curated-guide/.

are tracking the cases. We also see most lists are only tracking cases in the U.S.[12]
So, where might one go to keep track of what's going on? Let's look at a couple
of resources.

- **George Washington University:** GW has created a database of existing
  legislation related to AI and machine learning. [13] This searchable
  database includes more than *just* copyright cases but is a recommended
  database when researching AI litigation.   This is an example of a
  university effort to track the cases.

- **BakerHostetler**: They have an up-to-date Copyright and AI tracker,
  overseen by Theresa M Weisenberger.[14] This tracker allows the user to
  navigate new developments in AI copyright cases, as BakerHostetler
  monitors any status updates or key findings. This source is a wonderful
  option for those who do not know much about the impending litigation,
  as this case tracker includes case overviews and summaries of the key
  findings for any updates with each case. We found ourselves using this
  resource over and over to create this document.

- **Copyright Alliance:** This source has a section devoted to AI and
  Copyright.[15] They keep a table of the cases: where they are filed, the
  subject matter, and the date filed. [16] Each case then has a link to
  PacerMonitor. You can sign up for alerts from them, which keeps you
  up-to-date not only on the state of the cases but relevant AI news. They
  have also published a position paper on AI. Additionally, they have links
  to the positions their members have taken. This example of a non-profit
  advocacy group tracking the cases is a great source for those who want
  to delve into the conversations surrounding these major AI cases.

- **Wired Magazine:** This source tracks the cases, including a graphic of
  who is suing who.[17] They also list the status of the case. An example of
  a news organization tracking these cases, this site is easily digestible to

---

[12] *See, e.g.*, Aditya Kalra et al., *OpenAI faces new copyright case, from global book
publishers in India*, REUTERS (Jan. 24, 2025),
https://www.reuters.com/technology/artificial-intelligence/openai-faces-new-copyright-
case-global-publishers-india-2025-01-24.
[13] *DAIL - The Database of AI Litigation*, THE GEORGE WASHINGTON UNIVERSITY: ETHICAL
TECH INITIATIVE, https://blogs.gwu.edu/law-eti/ai-litigation-database/ (last visited Feb. 11,
2025).
[14] *Case Tracker: Artificial Intelligence, Copyrights and Class Actions*, BAKERHOSTELTER,
https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-
intelligence-copyrights-and-class-actions/ (last visited Feb. 11, 2025).
[15] Artificial Intelligence (*AI) and Copyright,* COPYRIGHT ALLIANCE,
https://copyrightalliance.org/education/artificial-intelligence-copyright/ (last visited
March 1, 2025).
[16] *Id.*
[17] Kate Knibbs, *Every AI Copyright Lawsuit in the US, Visualized*, Wired Magazine, (Dec.
19, 2024), https://www.wired.com/story/ai-copyright-case-tracker/.

the layperson. The inclusion of a visual case update section makes this source easily navigable.

● **McKool Smith**: Another law firm case tracker, this source solely focuses on updates on pending AI cases.[18] Instead of categorizing cases or cultivating a list of the most notable ones, this source follows major case developments through quick, weekly updates. This is a great source for anyone well-versed in current AI litigation but desires a site for consistent updates.

II. WHO ARE THE AI COMPANIES BEING SUED?[19]

We often see the same names as defendants. Who are they? How are they different? We have arranged the list in no particular order, but we did start with OpenAI, who continues to be at the center of the AI issues.

A.   OpenAI

OpenAI was founded in December 2015 after Sam Altman approached Elon Musk to discuss the potential dangers of artificial general intelligence ("AGI").[20] They then assembled a group of entrepreneurs and researchers in the technology field - including Greg Brockman, Ilya Sutskever, Trevor Blackwell, Vicki Cheung, Andrej Karpathy, Durk Kingma, John Schulman, Pamela Vagata, and Wojciech Zaremba - to start the non-profit AI research organization headquartered in San Francisco, California.[21]

OpenAI's charter states that their "mission is to ensure that artificial general intelligence – by which we mean highly autonomous systems that outperform humans at most economically value work–benefits all of humanity. We will

---

[18] *AI Litigation Tracker*, MCKOOL SMITH, https://www.mckoolsmith.com/newsroom-ailitigation (last visited March 1, 2025).

[19] Haley Pettingill, Tulane Law School, J.D., expected May 2025, and James Dearing, J.D., expected May 2025.

[20] AGI is a hypothetical artificial machine intelligence that could theoretically achieve human-level perception and cognition at inhuman speeds without biological constraints. Tim Mucci & Cole Stryker, *Getting Ready for Artificial Intelligence with Examples*, IBM.COM (Apr. 18, 2024), https://www.ibm.com/think/topics/artificial-general-intelligence-examples.

[21] Greg Brockman et. al. *Introducing OpenAI*, OPENAI.COM (Dec 11, 2015), https://openai.com/index/introducing-openai/. Due to possible conflicts of interest with Tesla's AI development and a fundamental disagreement on how the company should be operated, Elon Musk left in 2018. However, in February 2025, Musk made a $97 billion bid for the company, which has been denied. Keach Hagey, et. al., *The Inside Story of How Altman and Musk Went From Friends to Bitter Enemies*, Wall Street Journal (Feb. 14, 2025), https://www.wsj.com/tech/elon-musk-sam-altman-relationship-6889a77a.

attempt to directly build safe and beneficial AGI but will also consider our mission fulfilled if our work aids others to achieve this outcome."[22] The charter goes on to expound on the companies' four principles:

- broadly distributed benefits, with a "primary fiduciary duty…to humanity," explaining "We anticipate needing to marshal substantial resources to fulfill our mission, but will always diligently act to minimize conflicts of interest among our employees and stakeholders that could compromise broad benefit;"

- long-term safety, suggesting that for the sake of safety, if necessary they would "stop competing with and start assisting" with other projects that are "value-aligned" and "safety-conscious;"

- technical leadership, including being on the cutting edge of AI; and

- cooperative orientation with other research and policy institutions, including publishing their research, but "we expect that safety and security concerns will reduce or traditional publishing in the future…"[23]

Despite OpenAI's founding principles of openness, building safe AGI, and working for the good of the world, the company established a for-profit arm that would build outside capital through shareholders.[24] In 2019, Microsoft made a $1 billion investment in OpenAI (for the sole right to license OpenAI's technology for future products), which funded the company's expensive computing power.[25] This allowed OpenAI to develop numerous AI models, such as DALL-E, a generator that makes original works of art from text descriptions; Codex[26], a model that can understand and write code, and also to launch ChatGPT in 2022.[27]

ChatGPT is an AI chatbot that generates human-like dialogue from user questions (i.e., what is there to do in Akron, Ohio?) or commands (i.e., summarize this paragraph).[28] Deep learning models, such as ChatGPT, require existing training data to produce responses that replicate human cognition.[29] OpenAI states that while ChatGPT is not connected to the Internet, it was "trained on vast

---

[22]OpenAI Charter, https://openai.com/charter/, (last visited Feb. 13, 2025).

[23] *Id*.

[24] *Artificial: Episode 2, Selling Out*, THE JOURNAL, https://www.wsj.com/podcasts/the-journal/artificial-episode-2-selling-out/239740ec-071f-4493-96cb-91a6080e60ce (Dec. 10 2023) [hereinafter Episode 2].

[25] *Artificial: Episode 3, ChatGPT*, THE JOURNAL, https://www.wsj.com/podcasts/the-journal/artificial-episode-3-chatgpt/835fdc63-75ff-4ef1-b18c-f40537807cfe (Jan. 7 2024) [hereinafter Episode 3].

[26] Codex is most notable because it powers GitHub's AI tool Copilot. Karl Montevirgen, *OpenAI*, BRITANNICA MONEY (Feb. 11, 2025), https://www.britannica.com/money/OpenAI.

[27] *Id.*

[28] Episode 2, *supra* note 24.

[29] *Id*.

amounts of data from the internet written by humans."[30] OpenAI also operates with a "closed-source" system, meaning their codes, models, data sets, and algorithms are kept secret.[31] So, people began questioning where this training data came from and if OpenAI had the right to use it.[32] In 2023, OpenAI admitted to British Parliament that it would be "impossible" to train its AI without copyrighted material because "copyright today covers virtually every sort of human expression," and they argued that limiting the training data to public domain works "would not provide AI systems that meet the needs of today's citizens."[33] OpenAI maintains that it complies with the requirements of all applicable laws, including copyright laws, and that copyright law does not forbid its training.[34]

Between 2023 and 2024, OpenAI faced multiple lawsuits from the Authors Guild, media outlets, and independent authors and journalists whose work was allegedly used to train some of its products.[35] To combat these copyright infringement claims and the accompanying public scrutiny, OpenAI has pursued licensing deals with media companies and news organizations.[36] OpenAI has reportedly agreed to pay the most prominent publishers upwards of $10 million annually.[37] These partnerships allow OpenAI to train its language models on the publications' content (including paywalled information).[38] Publishers will get acknowledgment for their work through citations given by the chatbot and access

---

[30] *What is ChatGPT?*, OpenAI, https://help.openai.com/en/articles/6783457-what-is-chatgpt, (last visited Feb. 13, 2025).

[31] Rhiannon Williams & James O'Donnell, *We Finally Have a Definition for Open-Source AI*, MIT TECH. REVIEW (Aug. 22, 2024), https://www.technologyreview.com/2024/08/22/1097224/we-finally-have-a-definition-for-open-source-ai/.

[32] *Id.*

[33] OpenAI—Written Evidence: Hearing Before the House of Lords Communications and Digital Select Comm. Inquiry on Large language Models, LLM0113 (2023).

[34] *Id.*

[35] As of December 2024, 18 lawsuits have been filed involving OpenAI and its subsidiaries. According to Jonathan Gilham of Originality.ai these include: Canadian News Outlets v. OpenAI; Elon Musk v. Samuel Altman; Daily News LP v. Microsoft Corp.; The Intercept Media v. OpenAI, Raw Story Media, Inc. and Alternet Media Inc. v. OpenAI; The New York Times Co. v. OpenAI Inc.; Sancton v. OpenAI Inc.; Authors Guild v. OpenAI Inc.; Chabon v. OpenAI Inc; OpenAI Inc. v. Open Artificial Intelligence, Inc.; Doe 3 v. GitHub, Inc.; DOE 1 v. GitHub Inc., T. v. OpenAI LP; Walters v. OpenAI LLC; Silverman v. OpenAI Inc.; Tremblay v. OpenAI Inc; PM v. OpenAI LP.

[36] *OpenAI Licensing Deals with Legal Challenges While Securing Lucrative Agreements with Media Giants for AI Chatbot*, HYSCALER, (Jan. 12, 2024), https://hyscaler.com/insights/openai-licensing-deals-media-giants-ai/.

[37] Ivy Liu, *2024 In Review: A Timeline of The Major Deals Between Publishers and AI Companies*, DIGIDAY.COM (Dec. 27, 2024), https://digiday.com/media/2024-in-review-a-timeline-of-the-major-deals-between-publishers-and-ai-companies/.

[38] *Id.*

to the technology so they can build their own AI products and features.[39] While some news organizations are suing OpenAI, others are still weighing their legal options before entering negotiations, but as of January 2025, at least 14 have signed deals with OpenAI, including *Le Monde, Prisa, Financial Times,* Associated Press, Shutterstock, *The Atlantic,* News Corp., Condé Nast and Hearst.[40]

As of December 2024, there are 18 lawsuits involving OpenAI.[41] According to Jonathan Gilham of Originality.ai, the lawsuits include: *Canadian News Outlets v. OpenAI,*[42] *Elon Musk v. Samuel Altman, Gregory Brockman, and OpenAI,*[43] *Daily News Lp Et Al V. Microsoft Corporation,*[44] *The Intercept Media v. OpenAI and Microsoft,*[45] *Raw Story Media, Inc. and Alternet Media Inc. v. OpenAI,*[46] *The New York Times Company v. Openai Inc.,*[47] *Sancton v. OpenAI Inc.,*[48] *Authors Guild et al v. OpenAI Inc. et al,*[49] *Chabon v. OpenAI, Inc,*[50] *OpenAI, Inc. v. Open Artificial Intelligence, Inc.,*[51] *Doe 3 et al v. GitHub, Inc. et al,*[52] *DOE 1 et al v. GitHub, Inc. et a*l,[53] *T. et al v. OpenAI LP et al,*[54] *Walters v. OpenAI LLC,*[55] *Silverman, et al v. OpenAI Inc.,*[56] *Tremblay v. OpenAI Inc,*[57] and *PM et al v. OpenAI LP et al.*[58]

---

[39] *Id.*

[40]Charlotte Tobitt, *Who's Suing AI and Who's Signing: Major Deals with Google, OpenAI and Mistral Start Year*, PRESSGAZETTE.CO (Feb. 12, 2025), https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/.

[41]A Jonathan Gilham, "OpenAI and ChatGPT Lawsuit List," Originality.ai, December 17, 2024. This ist is expected bo be updated by the author.

[42] Canadian News Media Co. v. OpenAI, Inc*.*, No. CV-24-00732231 (Ont. Super. Ct. Just. Nov. 29, 2024).

[43] Musk v. Altman, No. 3:24-cv-04722 (N.D.Cal. Aug. 5, 2024).

[44] Daily News LP v. Microsoft Corp, No. 1:24-cv-03285 (S.D.N.Y. Apr. 30, 2024).

[45] The Intercept Media and Raw Story Media v. OpenAI, No*.* 1:24-cv-01515, (S.D.N.Y. Feb. 28, 2024).

[46] Raw Story Media, Inc. v. OpenAI, Inc., No. 1:24-cv-01514 (S.D.N.Y. Feb. 28, 2024).

[47] The N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 27, 2023).

[48] Sancton v. OpenAI, Inc., No. 1:23-cv-10211 (S.D.N.Y. Nov. 21, 2023).

[49] Authors Guild v. OpenAI Inc., No. 1:23-cv-08292 (S.D.N.Y. Sept. 19, 2023).

[50] Chabon v. OpenAI, Inc., No. 3:23-cv-04625 (N.D. Cal. Sept. 8, 2023).

[51] OpenAI, Inc. v. Open Artificial Intelligence, Inc., No. 4:23-cv-03918-YGR (9th Cir. Nov. 13, 2024).

[52] Doe 3 v. GitHub, Inc., No. 4:22-cv-07074 (N.D. Cal. Nov. 10, 2022).

[53] Doe 1 v. GitHub, Inc., No. 22-cv-06823-JST, 2024 WL 235217 (N.D. Cal. Nov. 3, 2022).

[54] T et al. v. Open LP, No. 3:23-cv-04557 (N.D. Cal. Sept. 5, 2023).

[55] Walters v. OpenAI, L.L.C., No. 1:23-cv-03122 (N.D. Ga. July 14, 2023).

[56] Silverman v. OpenAI, Inc., No. 3:23-cv-03416 (N.D. Cal. July 7, 2023).

[57] Tremblay v. OpenAI, Inc., 716 F. Supp. 3d 772 (N.D. Cal. 2024).

[58] PM v. OpenAI LP*,* No. 3:23-cv-03119 (N.D. Cal. June 28, 2023).

B. Stability AI

Emad Mostaque founded Stability AI in 2019 as a counter to the monopoly influence of big tech in the AI field.[59] The UK-based company develops open-source generative AI models[60], in contrast to OpenAI, Anthropic and Google's closed-source AI models.[61] With open-source models, researchers can study how the system works and users can "modify the system for any purpose—including to change its output—and [...] share it with others to use, with or without modifications."[62]

In 2022, Stability AI co-released its flagship product, Stable Diffusion, alongside researchers at LMU Munich and Runway AI.[63] Stable Diffusion produces "unique photorealistic images, [videos, and animations] from text and image prompts."[64] Stable Diffusion's user-friendly software saw considerable success with 150 million total downloads and 10 million users daily.[65] This boom allowed Stability AI to reach a $1 billion valuation just days after the product launched.[66] However, under Mostaques guidance, the once-promising startup had a hard time managing funds and attracting investors.[67] His idealistic views on how Stability AI would benefit humanity, such as "building bespoke AI models for nation states," proved to be impracticable.[68]

By early 2024, Stability AI cut 10% of its staff, and investors were furious with the mismanagement of operation expenses.[69] So, Mostaque resigned, and a new executive board was appointed with Prem Akkaraju as CEO and Sean Parker as Executive Chairman.[70] Under their leadership, the company has begun

---

[59] John E. Dunn, *Stability.ai Gets New CEO and Investment Dream Team to Start Rescue Rission*, CIO.com (Jun. 26, 2024), https://www.cio.com/article/2505518/stability-ai-gets-new-ceo-and-investment-dream-team-to-start-rescue-mission.html.

[60] Generative AI models create new and original data sets, like text, images or audio from user inputs. Altexsoft, *Generative AI Models Explained*, ALTEXSOFT (Sep. 4, 2024), https://www.altexsoft.com/blog/generative-ai/.

[61] Kenrick Cai & Iain Martin, How Stability AI's Founder Tanked His Billion-Dollar Startup, FORBES (Mar. 29, 2024), https://www.forbes.com/sites/kenrickcai/2024/03/29/how-stability-ais-founder-tanked-his-billion-dollar-startup/.

[62] Williams & O'Donnell, *supra* note 31.

[63] Stability.ai, *Celebrating One Year(ish) of Stable Diffusion ... and What a Year Itt's Been!*, STABILITY.AI (Oct. 3, 2023), https://stability.ai/news/celebrating-one-year-of-stable-diffusion.

[64] *What is Stable Diffusion*, AWS, https://aws.amazon.com/what-is/stable-diffusion/.

[65] Cai & Martin, *supra* note 61.

[66] Cai & Martin, *supra* note 61

[67] *Id.*

[68] *Id.*

[69] *Id.*

[70] *Id.*; STABILITY.AI, *supra* note 61.

restructuring efforts and attracted new investors[71] who have raised $80 million in funding.[72] The current increase in resources, signals a renewed confidence in Stability AI.[73] Further, the company's latest version of Stable Diffusion surpassed 2 million downloads during its first month of release in June 2024.[74] However, there are two U.S. lawsuits brought in 2023.[75]

A group of artists filed a class action lawsuit against Stability AI alleging that Stable Diffusion used their works as training images, and as a result it could "produce output images 'in the style' of those images."[76] Similarly, Getty Images filed a complaint asserting that Stability AI has not only copied more than 12 million photographs from the company, but that they "removed or altered Getty Images copyright management information, provided false copyright management information, and infringed Getty Images' famous trademarks."[77] Getty Images is seeking up to $150,000 for each infringed work, which could amount to $1.8 trillion.[78] Despite their legal battles, Stability AI's leaders continue to be committed to the company's founding principles, including making AI technology accessible to all.[79]

C. Anthropic PBC

Several former OpenAI employee's left the company after their partnership with Microsoft, and founded Anthropic PBC in 2021.[80] Anthropic is classed as a "public-benefit-company" – a corporation that makes a profit, but only by working on something that benefits humanity.[81] Since AI will have a significant impact on the world, Anthropic states that it is "dedicated to building systems that people can rely on and generating research about the opportunities and risks of

---

[71] Investors to include Greycroft, Coatue Management, Sound Ventures, Lightspeed Venture Partners, Sean Parker, Eric Schmidt and Prem Akkaraju. Kavout, *Stability AI's New Leadership and Funding: What it Means for the AI Startup*, KAVOUT SECTOR INSIGHTS (June 25, 2024), https://www.kavout.com/market-lens/stability-ais-new-leadership-and-funding-what-it-means-for-the-ai-startup.

[72] *Id.*

[73] *Id.*

[74] STABILITY.AI, *supra* note 63.

[75] *Andersen v. Stability AI Ltd.*, No. 23-cv-00201-WHO, 2024 U.S. Dist. LEXIS 143204 (N.D. Cal. Aug. 12, 2024); *Getty Images (US), Inc. v. Stability AI, Inc.*, 1:23-cv-00135-JLH; *see also* Cai & Martin, *supra* note 46.

[76] Anderson, *supra* note 59 at *3.

[77] Second Amended Complaint at 1, *Getty Images (US), Inc. v. Stability AI, Inc.*, 1:23-cv-00135-JLH.

[78] *Id.*

[79] Stability.ai, *supra* note 49.

[80] Andrew Andreev, *Anthropic PBC: History, Development, Products and Prospects,* API DRIVE: BLOG (Oct. 19, 2023), https://apix-drive.com/en/blog/useful/anthropic-pbc-history-development-products.

[81] Aaron Drapkin, *What is ClaudeAI and Anthropic? A Closer Look at ChatGPT's Rival*, TECH.CO (Jan. 2, 2024), https://tech.co/news/what-is-claude-ai-anthropic.

AI."[82] Although the founders of Anthropic left OpenAI because they feared Microsoft's contributions would make AI products too commercialized, they have agreed to partnerships with Google and Amazon.[83] Google's past deal with the company included a $2 billion investment, a 10% ownership in the company, and a large cloud contact between the two corporations. In January 2025, Google agreed to another $1 billion investment into the startup.[84] In 2024, Amazon increased its investments into the company to $4 billion, and Anthropic agreed to make Amazon Web Services ("AWS") its "primary cloud and training partner."[85]

With these large investments, Anthropic launched Claude, an AI assistant chatbot.[86] Like other chatbots, including ChatGPT, Claude relies on a reinforcement learning model to train its responses.[87] However, in an effort to reduce the undesired and harmful outputs of AI, Anthropic developed a new technique for AI called "Constitutional AI."[88] This method requires a developer to implement a set of principles and values (the "Constitution") which then creates the framework that the AI model must follow.[89] Instead of relying on an immense amount of human oversight to identify harmful outputs, Constitutional AI "encourages these models to evaluate their own outputs and refine themselves based on a predefined set of user-determined principles."[90]

Despite implementing extensive safety training for its AI software, Anthropic is still the subject of two copyright infringement lawsuits.[91] In order to replicate human-like cognition, Claude was trained, like all other AI chatbots, on vast amounts of data copied from the internet, including lyrics from innumerable musical works.[92] However, eight music publishers allege that Claude was not only improperly trained with their copyrighted song lyrics, but that Claude reproduced these song lyrics, without licensing, when prompted by users and for the

---

[82] *Company Homepage*, ANTHROPIC, https://www.anthropic.com/company (last visited Feb. 11, 2025).

[83] Hayden Field, *Google Agrees to New $1 Billion Investment in Anthropic*, CNBC (Jan. 22, 2025), https://www.cnbc.com/2025/01/22/google-agrees-to-new-1-billion-investment-in-anthropic.html.

[84] *Id*.

[85] *Id*.

[86] Drapkin, *supra* note 54.

[87] Riley Goodside & Spencer Papay, *Meet Claude: Anthropic's Rival to ChatGPT*, SCALE (Jan. 17, 2023), https://scale.com/blog/chatgpt-vs-claude.

[88] *Constitutional AI Explained*, TOLOKA (Apr. 19, 2024), https://toloka.ai/blog/constitutional-ai-explained/.

[89] *Id*.

[90] Neil Sahota, *Constitutional AI: Making AI Systems Uphold Human Values*, NEIL SAHOTA (Aug. 5, 2024), https://www.neilsahota.com/constitutional-ai-making-ai-systems-uphold-human-values/.

[91] Concord Music Grp., Inc. v. Anthropic PBC, 738 F. Supp. 3d 973 (M.D. Tenn. 2024); Andrea Bartz, et al. v. Anthropic PBC, No. 24-cv-05417, (N.D. Cal. 2024).

[92] *See generally Concord Music*, *supra* note 89.

commercial benefit of Anthropic.[93] Although Anthropic already had guardrails in place to make sure Claude did not informatively answer dangerous queries, (e.g., those "involving the production of chemical weapons,")[94] it has introduced more since the filing of the lawsuit.[95] In December 2024, Anthropic agreed to maintain its existing guardrails and to apply them against Claude reproducing the publisher's protected lyrics. In August 2024, a group of authors filed a similar class-action lawsuit alleging that Anthropic illegally downloaded, copied and generally misused their works to train Claude.[96]

D. GitHub

In 2008, GitHub was launched as an online, cloud-based collaboration website for software developers to share code, have a backup, and work on projects together.[97] The site acts as a Git server, which is an open-source version control[98] system.[99] Essentially, GitHub allows programmers to keep track of the changes to their code database, and co-programmers have access to this code and its history on their computers.[100] Further, the "version control system" allows developers to make changes to one aspect of a code, without compromising the main source code, and they can merge those changes into the finalized code database.[101] GitHub also offers a user-friendly interface that allows novices to use Git. [102] Almost every major software company, including Microsoft, Google, Twitter and Facebook, hosts their open source projects on GitHub.[103] In 2018, Microsoft acquired GitHub for $7.5 billion.[104] However, GitHub continues to operate as an independent entity, and still provides a platform for all levels of developers.[105]

---

[93] *Id.*

[94] *Constitutional Classifiers: Defending Against Universal Jailbreaks*, ANTHROPIC, (Feb. 3, 2025), https://www.anthropic.com/research/constitutional-classifiers.

[95] Blake Brittain, *Anthropic Reaches Deal on AI 'Guardrails' in Lawsuit Over Music Lyrics*, REUTERS (Jan. 3, 2025), https://www.reuters.com/legal/litigation/anthropic-reaches-deal-ai-guardrails-lawsuit-over-music-lyrics-2025-01-03/.

[96] Compl., Andrea Bartz, et al. v. Anthropic PBC, No. 24-cv-05417, (N.D. Cal. 2024).

[97] Jamie Juviler, *What is GitHub? (And What is it Used For?)*, HUBSPOT (Apr. 1, 2024), https://blog.hubspot.com/website/what-is-github-used-for.

[98] Most people have used some form of version control. For example, "Google Docs has a "Version History" tool where you can view changes to the document at different points in time." *Id.*

[99] *What is GitHub? A Beginner's Introduction to GitHub*, KINSTA (Nov. 17, 2023), https://kinsta.com/knowledgebase/what-is-github/.

[100] *Id.*

[101] *Id.*

[102] *Id.*

[103] Saurabh Mhatre, *The untold story of GitHub*, MEDIUM (Oct. 24, 2016), https://saurabhnativeblog.medium.com/the-untold-story-of-github-132840f72f56.

[104] *Microsoft to acquire Github for $7.5 Billion*, MICROSOFT NEWS CENTER (June 4, 2018), https://news.microsoft.com/2018/06/04/microsoft-to-acquire-github-for-7-5-billion/.

[105] *Id.*

Following the trend of integrating AI assistants into websites, GitHub introduced "Copilot" in 2021.[106] Copilot is an AI code-generator that suggests "lines of codes to programmers directly into their code editor."[107] According to GitHub, "Copilot has been trained on billions of lines of publicly-available code."[108] This caused concerns that GitHub might be reproducing line-by-line code from programmers who have open-source licenses on their platform.[109] So, in 2023, a class-action lawsuit for copyright infringement was filed against GitHub and Microsoft.[110] Github appears to be involved in only one lawsuit.[111]

E.   NVIDIA Corporation

Founded in 1993, NVIDIA specializes in producing graphics processing units (GPUs).[112] These are computer chips that "render graphics and images by performing rapid mathematical calculations."[113] Since GPUs are better at rendering images than traditional computer chips, they were originally associated with the video and computer game sector.[114] However, NVIDIA has since become the leading supplier of AI hardware and software because GPUs can perform the calculations concurrently – "making them more energy efficient and better able to handle sophisticated computing demands."[115] This efficient processing power is necessary for generative AI models, such as ChatGPT, which require thousands of GPUs to run.[116] Although other chip competitors, such as Intel, AMD and

---

[106] *Judge Throws Out Majority of Claims in GitHub Copilot Lawsuit*, LEGAL.IO (July 11, 2024), https://www.legal.io/articles/5516216/Judge-Throws-Out-Majority-of-Claims-in-GitHub-Copilot-Lawsuit.

[107] *Id.*

[108] *GitHub and Copilot Intellectual Property Litigation*, JOSEPH SAVERI LAW FIRM, https://www.saverilawfirm.com/our-cases/github-copilot-intellectual-property-litigation (last visited Feb. 16, 2025).

[109] LEGAL.IO, *supra* note 106.

[110] *Id.*

[111] Doe v. GitHub, Inc., 22-cv-06823-JST (N.D. Cal. May 11, 2023).

[112] Rob Wile, *Why Everyone is Talking About NVIDIA, the Nearly $3 Trillion-Dollar Company Fueling the AI Revolution*, NBC NEWS (Feb. 24, 2024), https://www.nbcnews.com/business/business-news/what-is-nvidia-what-do-they-make-ai-artificial-intelligence-rcna140171.

[113] Rahul Awati, *graphics processing unit (GPU) definition*, TECHTARGET, https://www.techtarget.com/searchvirtualdesktop/definition/GPU-graphics-processing-unit (last visited Feb. 16, 2025).

[114] Wile, *supra* note 112.

[115] *Id.*

[116] Sharon Goldman, *How NVIDIA Dominated AI - And Plans to Keep it That Way as Generative AI Explodes*, VENTUREBEAT (Feb. 23, 2023), https://venturebeat.com/ai/how-nvidia-dominated-ai-and-plans-to-keep-it-that-way-as-generative-ai-explodes/   (OpenAI reportedly used 10,000 Nvidia GPUs to train ChatGPT).

Google, have begun manufacturing their own GPUs, NVIDIA still has an advantage as the first-mover, holding about 88% of the GPU market.[117]

In September 2022, NVIDIA released "NeMo," a software program that simplifies the creation of conversational AI (i.e., chatbots, language translators, or voice assistants). NeMo offers developers an open-source toolkit with pre-trained large language models ("LLMs"). These LLMs generate human-like responses to user inputs, and, using NeMo, they can be integrated into various applications for numerous industries. By integrating NVIDIA's high-performance GPUs, their AI models have an accelerated training process where they "learn[] to predict the next word in a sentence based on the previous words. This process is repeated millions of times, allowing the model to learn linguistic patterns and nuances." However, like other AI models that utilize these deep learning techniques (i.e., ChatGPT and Claude), NeMo needs to collect and process vast amounts of data from books, articles, websites, and more. According to three authors, NeMo has been trained on datasets from "shadow library" websites that host vast quantities of unlicensed copyright material, including their books. Despite their copyright infringement complaint filed in 2024 and uncertainty about the future of AI, NVIDIA is still dominating the sector with its hardware and software. Prominent organizations, including AT&T, the University of Florida, Lowe's, and AWS, have adopted NeMo to bring generative AI into their industries.

F.   Microsoft

Bill Gates and Paul Allen founded Microsoft on April 4, 1975.[118] The company initially focused on selling microcomputer software before expanding into operating systems.[119] Microsoft Office has become one of the most widely used software suites, featuring popular tools such as Word, Excel, and PowerPoint.[120] Over the years, Microsoft has extended its market influence by acquiring and investing in major technology companies, including LinkedIn, GitHub, Skype, and OpenAI.[121] It also maintains a strong presence in the gaming industry as the owner of the Xbox console.[122] In 2024, Fortune ranked Microsoft as the 26th largest company in the world according to revenue.[123] As of December 27, 2023, Microsoft had pledged $13 billion to OpenAI and announced plans to integrate OpenAI's technology into the Bing search engine.[124]

---

[117] *Id.*

[118] Mark Hall, *Microsoft Corporation*, Britannica Money (Feb. 12, 2025), https://www.britannica.com/money/Microsoft-Corporation.

[119] *Id.*

[120] *Id.*

[121] *Id.*

[122] *Id.*

[123] Scott DeCarlo, *Fortune Global 500*, Fortune, https://fortune.com/ranking/global500/.

[124] Britannica Money, *supra* note 118.

G.   Google

Google was founded in 1998 by Sergey Brin and Larry Page.[125] The company operates the world's most widely used search engine, handling more than 70% of global online search requests.[126] Since 2015, Google has been a subsidiary of the holding company Alphabet Inc.[127] In 2023, Alphabet generated $175 billion in revenue from advertising based on users' search activity.[128] In 2024, Fortune ranked Alphabet as the 17th largest company in the world by revenue.[129] Today, Google offers far more than just search services, providing over 50 additional online tools, including Gmail and its suite of document creation applications, such as Google Docs and Google Sheets.[130]

In April 2023, Google CEO Sundar Pichai introduced Bard, the company's first conversational AI model.[131] Initially launched as an experimental service, Bard received widespread criticism.[132] In December 2023, Google released an updated version, Google Gemini, which has faced its own controversies, particularly regarding issues with inappropriate image generation.[133] Google Gemini for all intents and purposes is the same thing as Bard. Beyond Google Gemini, the company began the first publicly-discussed explorations of deep learning AI models with the Google Brain project in 2011, and the acquisition of the UK based deep learning startup "Deepmind" in 2014.[134] Since then, Google has championed the use of deep learning AI models by integrating them into many of their services.[135] These deep learning models "attempt to mimic the way [human] brains sort and process information" by building complex data processing configurations, called artificial "neural networks."[136] Google uses these deep neural networks to create image recognition applications, such as Google Photos, which allows users "to search photos by what's in them."[137] Similarly, the processing technology has accelerated natural language

---

[125] *Id.*

[126] *Id.*

[127] *Id.*

[128] *Id.*

[129] FORTUNE, *supra* note 9.

[130] BRITANNICA MONEY, *supra* note 86.

[131] Frannie Comstock, *Google Gemini*, BRITANNICA TECHNOLOGY (Feb. 11, 2025), https://www.britannica.com/technology/Google-Gemini.

[132] *Id.*

[133] *Id.*

[134] Berrnard Marr, *The Amazing Ways Google Uses Deep Learning AI*, Forbes (Aug. 8, 2017), https://www.forbes.com/sites/bernardmarr/2017/08/08/the-amazing-ways-how-google-uses-deep-learning-ai/.

[135] *Id.*

[136] *Id.*

[137] Marian Croak & Jeff Dean, *A Decade in Deep Learning, and What's Next*, GOOGLE BLOG (Nov. 18, 2021), https://blog.google/technology/ai/decade-deep-learning-and-whats-next/.

generation.[138] In 2016, deep learning was implemented into Google Translate which increased its quality and reduced error rates.[139] Further, in the past decade, Google developed deep learning models that can read house numbers and street signs to better help users find addresses in Google Maps, that allow users to interact with the world through Google Lens, that give better YouTube recommendations, that can predicate flood forecasting, that can detect diabetic retinopathy and cancer cells, and more. [140] Even Google's self-driving car division, Waymo, uses deep learning models to better analyze and react to their surroundings. [141] Most recently, in 2024, Google started using Gemini's technology to give summarized "AI overviews" of a search result before the traditional web page links.[142]

H. Bloomberg

Michael Bloomberg founded Bloomberg L.P. in 1981 after leaving Salomon Brothers. [143] The company initially focused on developing the Bloomberg Terminal, a system that provided real-time market data, financial calculations, and analytics to Wall Street firms.[144] Over time, Bloomberg expanded its services to include Bloomberg News, Bloomberg Television, Bloomberg Radio, and publications like Bloomberg Businessweek. These platforms deliver comprehensive business and financial information to a global audience. [145]Bloomberg has also grown through strategic acquisitions, such as the purchase of BusinessWeek magazine in 2009, which was subsequently renamed Bloomberg Businessweek. [146] In 2023, the company acquired Broadway Technology, a provider of high-performance front-office financial technology solutions.[147]

In recent years, Bloomberg has increased its focus on artificial intelligence (AI).[148] The company launched AI-powered news summaries for investors, where

---

[138] *Id.*

[139] *Id.*

[140] *Id.*

[141] Marr, *supra* note 134.

[142] Dan Milmo & Nick Robins-Early, *Google Rolls Out AI-Generated, Summarized Search Results in US,* THE GUARDIAN (May 14, 2024), https://www.theguardian.com/technology/article/2024/may/14/google-ai-search-results.

[143] *Mike's Story*, MIKE BLOOMBERG, https://www.mikebloomberg.com/about/ (last visited Feb. 16, 2025).

[144] Mara Lesemann, *Here's Why Bloomberg Stock Doesn't Exist*, INVESTOPEDIA (Jan. 2, 2025), https://www.investopedia.com/articles/markets/121715/bloomberg-stock-doesnt-exist-here-why.asp.

[145] *Id.*

[146] *Id.*

[147] *Bloomberg Completes Acquisition of Broadway Technology*, BLOOMBERG (Oct. 3, 2023), https://www.bloomberg.com/company/press/bloomberg-completes-acquisition-of-broadway-technology/.

[148] *Id.*

generative AI creates the summaries, which are then reviewed by analysts.[149] Additionally, Bloomberg developed the Bloomberg Artificial Intelligence Index (BAIAT), which is both representative and diversified across key segments of the entire AI ecosystem.[150]

In March 2024, Bloomberg L.P. faced a lawsuit filed by authors, including former Arkansas Governor Mike Huckabee and best-selling Christian writer Lysa TerKeurst. The plaintiffs alleged that Bloomberg used their copyrighted books without permission to train its large language model, BloombergGPT, constituting copyright infringement. In response, Bloomberg filed a motion to dismiss the lawsuit, asserting that its use of the authors' works was protected under the fair use doctrine. The company argued that the materials were utilized as part of internal research into generative AI capabilities and were not employed for commercial purposes.[151]

I.   Databricks and Mosaic ML

Databricks was founded in 2013 by Ali Ghodsi, Ion Stoica, Matei Zaharia, Patrick Wendell, Reynold Xin, Andy Konwinski, and Arsalan Tavakoli-Shiraji, the original creators of Apache Spark.[152] The company initially focused on developing a unified analytics platform to streamline big data processing and machine learning workflows.[153] Over the years, Databricks has expanded its offerings to include the Lakehouse Platform, which combines the capabilities of data lakes and data warehouses, enabling organizations to manage and analyze vast amounts of data efficiently.[154] This platform supports various data workloads, including business intelligence, data engineering, and machine learning.[155]

Databricks has also grown through strategic acquisitions. In June 2020, the company acquired Redash, an open-source tool for data visualization and building

---

[149] *Exame: Generative AI Sparks Enthusiasm but Still has Limitations in the Financial Sector, Says Bloomberg Executive*, BLOOMBERG (Jan. 31, 2025), https://www.bloomberg.com/company/press/exame-generative-ai-sparks-enthusiasm-but-still-has-limitations-in-the-financial-sector-says-bloomberg-executive.

[150] *Indexing the AI theme: From "pure play" to the "show me" narrative*, BLOOMBERG (Nov. 27, 2024), https://www.bloomberg.com/professional/insights/artificial-intelligence/indexing-the-ai-theme-from-pure-play-to-the-show-me-narrative/

[151] Blake Brittain, *Bloomberg asks US court to toss copyright lawsuit over AI training*, REUTERS (Mar. 25, 2024), https://www.reuters.com/legal/litigation/bloomberg-asks-us-court-toss-copyright-lawsuit-over-ai-training-2024-03-25/.

[152] *Leadership*, DATABRICKS, https://www.databricks.com/company/founders (last visited Feb. 16, 2025).

[153] *Id.*

[154] *Id.*

[155] *Id.*

interactive dashboards.[156] In 2021, it purchased 8080 Labs, a German no-code company known for its data exploration tool, bamboolib[157] . In June 2023, Databricks acquired MosaicML, a generative AI platform, to enhance its AI capabilities.[158] In 2024, Databricks introduced the Databricks Data Intelligence Platform, integrating MosaicML's generative AI technology to help customers better understand and utilize their proprietary data. This move underscores the company's commitment to advancing AI and machine learning solutions.[159]

In March 2024, Databricks, Inc. and its subsidiary MosaicML faced a class-action lawsuit filed by authors Abdi Nazemian, Brian Keene, and Stewart O'Nan.[160] The plaintiffs allege that Databricks' MosaicML models were trained on datasets containing nearly 197,000 books sourced from shadow libraries like Bibliotik, which host unlicensed copyrighted material.[161] The authors claim their works were used without consent, credit, or compensation, constituting direct copyright infringement.[162] According to Wired Magazine, Databricks and Mosaic ML are involved in one case, where NVIDIA is also involved.[163]

J.   ROSS Intelligence (ROSS)

This was the first AI case filed, and had a decision on revised decision on Summary Judgement as of February 11, 2025.[164] Coming out of the University of Toronto, ROSS Intelligence described itself as building "AI-driven products to augment lawyers' cognitive abilities."[165] Their "Our Story" explains that in 2014 two computer scientists and a lawyer decided to make legal services more accessible.[166] "Over a short period, ROSS raised millions in funding from Silicon Valley's most prominent investors and received top recognitions from Forbes and the American Bar Association Journal." The service included question-based search, find similar language, document analyzer, question-focused overviews,

---

[156] Reynold Xin, *Welcoming Redash to Databricks*, DATABRICKS (June 24, 2020), https://www.databricks.com/blog/2020/06/24/welcoming-redash-to-databricks.html.

[157] For more, *see* https://docs.databricks.com/aws/en/archive/notebooks/bamboolib.

[158] *Databricks Completes Acquisition of MosaicML*, DATABRICKS (July 19, 2023), https://www.databricks.com/company/newsroom/press-releases/databricks-completes-acquisition-mosaicml.

[159]*Id*.

[160] Stewart O'Nan et al. v. Databricks, Inc. et al., No 3:24-cv-01451 (N.D.Cal. 2024).

[161] *Id*.

[162] Databricks, Inc. Large Language Model Litigation, JOSEPH SAVERI LAW FIRM, https://www.saverilawfirm.com/databricks-inc.-large-language-model-litigation   (last visited Feb. 16, 2025). .

[163]Knibbs, *supra* note 9.

[164] Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-CV-613-SB (D. Del.)

[165] *About Us*, ROSS INTELLIGENCE, https://www.rossintelligence.com/about-us (last visited March 1, 2025)

[166] *Id.*

and case treatments. They were charging $69-89 a month.[167] And they used, through a third party, Thomson Reuter's headnotes for training their AI system. By January 31, 2021 ROSS was no longer in business. But that's not what their website reports. Instead, the "Enough" posts claims that they are filing antitrust claims against Thomson Reuters.[168]

K.   *Perplexity*

Perplexity is an AI search engine, which among other things, lists the sources for the responses and includes footnotes/references to the output. They are also licensing with companies to use their materials. Perplexity is quite new, founded in 2022, with founders coming from OpenAI, Meta and Quora. You can search videos and also create images. In short, they are billing themselves as the "accurate" AI tool.[169] By December 2024, Perplexity AI had a valuation of $9 billion, and had raised in its fourth round of funding $500 million.[170] They had started the year as valued at $500 million.[171] They have debuted a revenue sharking model for publishers, and also have licensing deals with man. The company is being sued by News Corp. and the New York Post.

L. *AI voice generators*

Lovo is an AI voice generation company that was recently sued for illegal cloning of voice actors' voices. Who is Lovo as a company? "Our award-winning voice generator and text to speech software is packed with 500+ voices in 100 languages."[172] Voice cloning technology replicates an existing yuan voice.[173] Udio and Suno are also getting suede. These are both prominent AI music platforms that allow users to generate songs from text prompts. They are also getting sued by all three major music labels. Finally, ElevenLabs is also a text to speech AI generator.

---

[167] *Pricing*, ROSS INTELLIGENCE, https://www.rossintelligence.com/pricing, (last visited March 1, 2025).

[168] The Founders, *Enough,* ROSS NEWS, https://blog.rossintelligence.com/post/enough, January 25, 2021.(last visited March 2, 2025).

[169] Perplexity, http://www.perplexity.ai (last visited March 1, 2025).

[170] Hayden Field, Perplexity AI in final stages of raising $500 million round at $9 billion valuation, November 5, 2025,
https://www.cnbc.com/2024/11/05/perplexity-ai-nears-500-million-funding-round-at-9-billion-valuation.html (last visited March 1, 2025)

[171] *Id.*

[172] *AI Voice Generator: Most Realistic AI Text to Speech,* LOVO AI, https://lovo.ai/ (last visited March 2, 2025).

[173] James G. Gatto, *Lovo "Voices" Opposition to Suit Over "Kitchen-Sink" Approach to Alleged AI Voice Cloning*, NAT. L. REV., ,https://natlawreview.com/article/lovo-voices-opposition-suit-over-kitchen-sink-approach-alleged-ai-voice-cloning (March 1, 2025).

III. THE AI COPYRIGHT CASES THEMSELVES: SOME SUMMARIES

This is an illustrative – but not exhaustive – exploration of the major AI cases as of February 2025. As discussed above, what counts as a case is unclear, and most lists do not include cases outside of the U.S. We have only reviewed select U.S. cases in this study.

A. *Alter v. OpenAI Inc., No. 1:23-cv-10211, (S.D.N.Y. Feb. 6, 2024)*[174]

Alter v. OpenAI is the product of three consolidated cases filed against Microsoft and OpenAI on behalf of various journalists and authors in the Southern District of New York.[175] The consolidated complaint alleges that Microsoft and OpenAI infringed on the plaintiff's copyrighted works when training OpenAI's model without their permission.[176] OpenAI is a research organization founded in 2015 by various researchers and entrepreneurs such as Sam Altman, Peter Thiel, and Elon Musk.[177] OpenAI runs and operates ChatGPT, an AI chatbot that provides generated responses based on user prompts.[178] ChatGPT's AI models are referred to as "GPT-N." Pertinent to this litigation is the GPT-3 model released in 2020, the GPT-3.5 model released in 2022, and the GPT-4 model released in 2023.[179] Microsoft first invested in OpenAI in 2019 with a $1 billion investment in the company, but the explosion in popularity in AI from the release of ChatGPT caused Microsoft to invest another $10 billion in early 2023.[180]

To understand the merits of the case, it is important to generally understand how ChatGPT works and how it uses copyrighted material. One of ChatGPT's innovations is its use of "unsupervised learning," where the program is given

---

[174] Gerardo Gomez, Tulane Law School, J.D. expected May 2026.

[175] The three individual cases were *Authors Guild v. OpenAI Inc.*, *Alter v. OpenAI Inc.*, and *Basbanes v. Microsoft Corporation*. The three cases included claims for direct copyright infringement and vicarious infringement, while *Alter v. OpenAI Inc.* did not include a count for contributory infringement. *See generally Alter v. OpenAI*, BAKERHOSTETLER, https://www.bakerlaw.com/alter-v-openai/ (last visited Feb. 11, 2025).

[176] *See id.*

[177] OpenAI consists of two principal entities: OpenAI Inc. - which is the research-focused non-profit - and OpenAI Global LLC - which is a for-profit entity for the commercialization of OpenAI products. *See* Karl Montevirgen, *OpenAI*, BRITANNICA (Feb. 14, 2025), https://www.britannica.com/money/OpenAI.

[178] *See id.*

[179] Compl., ¶ 84, 85, 89, Alter v. OpenAI, No. 1:23-cv-10211, (S.D.N.Y., Feb. 6, 2024).

[180] [5] Alexandra Garfinkle, *Microsoft's near-term fate is in OpenAI's hands — for better or worse*, YAHOO!FINANCE (Dec. 20, 2023), https://finance.yahoo.com/news/microsofts-near-term-fate-is-in-openais-hands--for-better-or-worse-205250812.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAAWTNbykeKPUs2Idjd2kNEiE6NfPV4sosY8sJNS6H6g6GqeDrcs3pwGsLq0D6k_KCIj0VRjMdlIDhwc2zZdY_b3NNiVre-4OJvPW2yXG8c-T3C7z8-Ck5Jom3aB4ixuR8gd-gSgZwCXgnyRTm8AROZL5g_epehmxr8gAB5quOijL.

broad parameters and fed large amounts of data.[181] The model is unsupervised because it processes all of this data, which is equivalent to almost all available data on the open internet, to generate a knowledge basis.[182] Different models have introduced new kinds of data that can be used for training, such as GPT-4o which allows models to be trained on audio and image data.[183]

The first complaint was filed by the Authors Guild alongside several acclaimed authors like George R. R. Martin on September 19, 2023.[184] This complaint alleged that the defendants had trained ChatGPT on the plaintiff's books without their permission.[185] The Authors Guild plaintiffs filed three causes of action: direct copyright infringement by OpenAI LP for reproducing the plaintiff's works, vicarious liability from OpenAI Inc. and its international entity, and contributory copyright infringement on OpenAI's other subsidiaries for assisting, controlling, and managing on the AI model's training on protected material.[186] On November 21, 2023, the Alter plaintiffs filed their own complaint, alleging only the direct and contributory copyright infringement claims.[187] Lastly, the Basbanes plaintiffs filed their complaint, alleging the same causes of action as the Authors Guild plaintiffs, on January 5, 2024.[188] The Basbanes plaintiffs filed to consolidate their case with the Authors Guild on January 23, 2024, while the Alter plaintiffs finalized consolidating their case the following day with the Authors guild.[189] The litigation is currently ongoing, and the court set January 8, 2025, as the deadline to submit the final amended pleadings.[190]

In the consolidated complaint, the plaintiffs allege that to train GPT-3, GPT-3.5 and GPT-4 ("GPT Models"), OpenAI had reproduced copyrighted books without the author's consent. OpenAI admits that the GPT Models were trained using "large, publicly available datasets that include copyrighted works."[191] Furthermore, OpenAI admitted that analyzing the large amounts of data necessarily involved making reproductions of the training data.[192] Had OpenAI not used the copyrighted books to train the GPT Models, their quality would have been significantly reduced.[193] The plaintiffs allege that OpenAI knew their

---

[181] Harry Guinness, *How does ChatGPT work?*, ZAPIER (Nov. 18, 2024), https://zapier.com/blog/how-does-chatgpt-work/.

[182] *See id.*

[183] *See id.*

[184] BAKERHOSTETLER, *supra* note 124.

[185] *Id.*

[186] *Id.*

[187] *Id.*

[188] *Id.*

[189] *Id.*

[190] *Id.*

[191] Compl. ¶ 99, Alter v. OpenAI, No. 1:23-cv-10211, (S.D.N.Y., Feb. 6, 2024).

[192] *See id.* at ¶ 100.

[193] *See id.* at ¶ 101.

training material included copyrighted material, yet continued to use it to train the GPT Models without obtaining permission from the authors.[194]

The consolidated complaint further discusses Microsoft's vital role in training the GPT Models. Microsoft's supercomputer, Azure, was specifically built to aid OpenAI train the GPT Models.[195] Thus, while OpenAI designed the AI models themselves, Microsoft built the computer system that enables the reproduction of copyrighted material.[196] Furthermore, Microsoft staff worked closely with OpenAI employees during the training process.[197] Therefore, the plaintiffs allege that Microsoft knew of OpenAI's copyright infringement and contributed to it by providing the sophisticated hardware necessary to train the GPT Models.[198]

In their answer to the plaintiff's complaint, OpenAI argues that their use of the copyrighted books constitutes paradigmatic transformative fair use.[199] OpenAI claims that their model just extracted information about the work rather than replicating the expression itself.[200] Furthermore, OpenAI argues that the GPT Models generate brand new material, using the copyrighted works to obtain an understanding of language, reasoning, and the world.[201] Similarly, while Microsoft admits that it collaborated with OpenAI to design the Azure supercomputer to train large language models, it denies knowingly infringing on copyrighted material.[202]

Tags: copyright infringement, training AI Models, OpenAI, Microsoft, fair use

B. *Andersen v. Stability AI Ltd., No. 23-cv-00201-WHO, (N.D. Cal.).* [203]

In *Anderson v. Stability AI*, the plaintiffs allege copyright infringement, along with inducement, DMCA violations, false endorsement, and trade dress infringement. In October 2023, artists Sarah Andersen, Kelly McKernan, and Karla Ortiz filed a class action lawsuit on behalf of themselves and other artists in the United States District Court for the Northern District of California, against defendants Stability AI ("Stability"), Midjourney, Inc. ("Midjourney"), and DeviantArt, Inc. ("DeviantArt").[204] The plaintiffs alleged direct copyright infringement under 17 U.S.C. § 106, violation of the Digital Millennium

---

[194] *See id.* at ¶ 128.
[195] *See id.* at ¶ 133.
[196] *See id.* at ¶ 135-36.
[197] *See id.* at ¶ 136.
[198] *See id.* at ¶ 137.
[199] OpenAI Defendants' Answer to First Consolidated Class Action Complaint at 3, *Alter v. OpenAI*, No. 1:23-cv-08292-SHS, at *3 (S.D.N.Y. Mar. 4, 2024).
[200] *Id.*
[201] *Id.*
[202] *See Alter*, No. 1:23-cv-08292-SHS at *2.
[203] Marie Gosserand, Tulane University Law School Expected Graduate May 2025.
[204] Andersen v. Stability AI Ltd., 700 F. Supp.3d 853 (N.D. Cal. 2023).

Copyright Act under 17 U.S.C. §§ 1201-1205 ("DMCA"), violation of the Right to Publicity under California Civil Code § 3344, violation of the Common Law Right of Publicity, Unfair Competition under California's Business and Professional Code § 17200, and declaratory relief against all defendants, in addition to a breach of contract claim against DeviantArt alone.[205] The claims challenged the Defendants' creation and/or use of generative text-to-image artificial intelligence (AI) software called Stable Diffusion, alleging that their copyrighted works were used in the creation and training of the software.[206]

Stable Diffusion is a generative image AI diffusion model that creates high-quality images based on a text prompt.[207] A diffusion model is a type of generative AI model whose function is inspired by the process of thermodynamic diffusion.[208] Similarly to how a drop of food coloring spreads in water to create a uniform color, diffusion models take an image and diffuse it by altering the pixels until the original image becomes "TV static."[209] By doing so, the model is learning how to reverse the process to take the static-y image and diffuse it backwards into an image.[210] The basis for this process is that it is very easy for computers to generate TV static and then use the randomness of the generated static to produce different images each time, even if the same prompt is used.[211]

Stable Diffusion was first built in 2021 through a collaboration between Runway AI ("Runway"), a research company specializing in AI generated art, and researchers at the University of Munich.[212] Stability, known as a leading company in the field of open-source generative AI, later joined the project to help pay the computing costs associated with training the model on large data sets.[213] Stability presented Stable Diffusion as a product in 2022 under a permission source

---

[205] *Id.*

[206] Savan Dhameliya, *US Court allows claims against text-to-image AI Companies: Sarah Andersen v. Stability AI*, IPRMENTLAW (Aug. 25, 2024), https://iprmentlaw.com/2024/08/25/us-court-allows-claims-against-text-to-image-ai-companies-sarah-anderson-v-stability-ai/.

[207] *Artificial Intelligence for Image Research*, UNIVERSITY OF TORONTO LIBRARIES (Feb. 7, 2025), https://guides.library.utoronto.ca/image-gen-ai.

[208] *Dave Bergmann and Cole Stryker, What are diffusion models?,* IBM (Aug. 21, 2024), https://www.ibm.com/think/topics/diffusion-models.

[209] *Artificial Intelligence for Image Research*, *supra* note 205.

[210] *Id.*

[211] *Id.*

[212] Will Douglas Heaven, *The original startup behind Stable Diffusion has launched a generative AI for video*, MIT TECH. REV (Feb, 6, 2023), https://www.technologyreview.com/2023/02/06/1067897/runway-stable-diffusion-gen-1-generative-ai-for-video/.

[213] *Id.*; *Stability AI*, DELPHI DIGITAL, https://members.delphidigital.io/projects/stability-ai (last visited Feb. 11, 2025).

license[214], which has now become one of the most well-known AI-powered text-to-image diffusion models today.[215]

The claims implicated each of the Defendants in some way for their association with the Stable Diffusion program. In training Stable Diffusion, Stability allegedly used datasets known as LAION-5B and LAION-400M, containing billions of images taken from the internet, including the plaintiffs' works.[216] This training is what enables software to generate new images that replicate or closely resemble the style of the images in the datasets they were trained with.[217] The Stable Diffusion product "is alleged to be a 'software library' providing 'image-generating services' to products produced and maintained" by Defendants Midjourney and DeviantArt.[218]

DeviantArt, founded in 2000, is known as an online community for digital artists to post and share their work.[219] In November 2022, DeviantArt released its commercial product "DreamUp" to its paying customers, which utilizes Stable Diffusion to produce images. Plaintiffs claim that one of the LAION datasets used in training Stable Diffusion was comprised of images from websites, including DeviantArt's own site.[220] Thus, by incorporating Stable Diffusion on its site through DreamUp, plaintiffs alleged that DeviantArt violated its own terms of service by using content for commercial purposes and without consent, in addition to violating its privacy policy.[221]

Midjourney, Inc., is an independent research lab in San Francisco, California.[222] Midjourney created the Midjourney Product, released in July 2022, and is alleged to be a commercial product producing images from text prompts similarly to Deviant Art's product "DreamUp."[223] Plaintiffs allege that the Midjourney Product uses Stable Diffusion and also that it was trained on a subset of images used to train Stable Diffusion.[224]

In an order on October 30, 2023, Judge Orrick granted plaintiffs leave to amend their complaint that was "defective in numerous aspects," ordering Plaintiffs to "provide clarity regarding their theories of how each defendant separately violated their copyrights, removed or altered their copyright management information, or violated their rights of publicity and plausible facts in support."[225]

---

[214] *Stability AI supra* note 213.

[215] *Id.*

[216] *Andersen*, 700 F. Supp. 3d at 860-862.

[217] *Id.* at 861.

[218] *Id.* at 860.

[219] *Id.* at 861.

[220] *Id.*

[221] *Id.*

[222] *Id.*

[223] *Id.*

[224] *Andersen v. Stability AI*, BAKERHOSTETLER, https://www.bakerlaw.com/andersen-v-stability-ai/ (last visited Feb. 11, 2025).

[225] Order on Mot. To Dismiss and Strike at 1, *Andersen v. Stability AI*, No. 3:23-cv-00201-WHO, (N.D. Cal. Oct. 30, 2023).

Plaintiffs' first amended complaint, filed on November 29, 2023, included more Plaintiffs and added Runway AI, the original collaborator on the Stable Diffusion Project, as a defendant.[226] On August 12, 2024, the court issued an order granting in part and denying in part the Defendants' motions to dismiss the first amended complaint.[227] Regarding the direct and induced copyright infringement claims, the court denied each defendant's motion to dismiss.[228] Next, the court dismissed the DMCA claims with prejudice.[229] It reasoned that Plaintiffs' § 1202(a) claims against Stability and Runway were insufficient because it was "'implausible' that readers of the license disclosures for Stable Diffusion, under which copyright ownership was asserted for the model, 'would understand that defendants are claiming rights to or conveying any false information regarding the rights of plaintiffs.'"[230] Regarding Plaintiffs' claims under 1202(b) against Stability, Runway, and Midjourney, Judge Orrick granted the motion to dismiss with prejudice on the grounds that a claim involving the removal of copyright management information requires identicality between the original work and the copy (per *GitHub*), which the plaintiffs failed to establish.[231]

The court dismissed the plaintiffs' unjust enrichment claims against the defendants but granted leave to amend on the grounds the claims were preempted by the Copyright Act because they were tied to the defendants' alleged use of the plaintiffs' copyrighted works in training its AI models without consent and plaintiffs had not identified an extra element that "change[d] the nature of those state law claims to protect something other than rights protected under the Copyright Act."[232] Lastly, the court found the false endorsement and vicarious trade dress infringement claim against Midjourney sufficient to survive a motion to dismiss, but dismissed the breach of contract claim against DeviantArt with prejudice on the grounds that DeviantArt itself played no role in the scraping or training of Stable Diffusion.[233]

The plaintiffs filed a second amended complaint on October 31, 2024.[234] Against Stability, the plaintiffs alleged 1) direct copyright infringement in the LAION-5B dataset by downloading copies of copyrighted works from the URLs in the dataset, using the copies to train its AI models, and distributing the AI models to the public; 2) induced copyright infringement "by materially

---

[226] Andersen v. Stability AI, No. 3:23-cv-00201-WHO, First Amended Complaint (N.D. Cal. Nov. 29, 2023).

[227] Order Granting In Part and Denying In Part Mots. To Dismiss First Amended Complaint, Andersen v. Stability AI, No. 3:23-cv-00201-WHO, (N.D. Cal. Aug. 12, 2024).

[228] *Id.*

[229] *Id.*

[230] *Id.*

[231] *Id.*

[232] *Id.* at 14.

[233] *Id.*

[234] Andersen v. Stability AI, No. 3:23-cv-00201-WHO, Seconded Amended Complaint (N.D. Cal. Oct. 31, 2024).

contributing to downstream copyright infringement through distributing the allegedly infringing … AI models for free under the MIT license; and 3) and violations of DMCA § 1202 by falsely asserting copyright in its allegedly infringing AI models by distributing them under the MIT license, and knowingly removing or altering CMI from copies of the plaintiffs' works in training its AI models, which caused its AI models to generate images based on plaintiffs' copyrighted works with CMI removed or altered. [235] Against Midjourney, plaintiffs alleged 1) direct copyright infringement of their works in the LAION-400M and LAION-5B dataset by downloading copies of the plaintiffs' works and using the copies to train its AI models and distributing the AI model to the public; 2) violations of DMCA § 1202(b) by knowingly removing or altering CMI from the copies of plaintiffs' works when training its AI; 3) false endorsement by using plaintiff's names commercially in advertising its AI product; and 4) vicarious trade dress infringement by encouraging and inducing users of its AI product to input prompts with artists' names and generate images "featuring the recurring visual elements and artistic techniques constituting plaintiff's protectable trade dress."[236] Against Runway, plaintiffs alleged 1) direct copyright infringement in using their copyrighted works in the LAION-5B dataset and using the copies to train their AI models and distributing the model to the public; 2) induced copyright infringement by materially contributing to downstream copyright infringement by distributing the infringing Stable Diffusion model for free under the "Creative ML Open RAIL-M license"; and 3) violations of DMCA § 1202 by falsely asserting that third parties have a copyright in the Stable Diffusion model when distributing them under the Creative ML RAIL-M license in violation of §1202(a), and knowingly removing or altering CMI from copies of the plaintiffs' works in training its AI models, which caused its AI models to generate images based on plaintiffs' copyrighted works with CMI removed or altered in violation of §1202(b). [237] Lastly, against DeviantArt, plaintiffs alleged direct copyright infringement of their copyrighted works in the LAION-5B dataset by incorporating the Stable Diffusion model into its product "DreamUp" and distributing that model to the public.[238] The trail for this case is set for April 5, 2027.[239]

The original dismissal of all Copyright Management Information claims in this case exhibits a problem all litigants are facing in AI infringement cases asserting Section 1201 claims.[240] First, the standard set out by 1202(a) is very difficult to prove without significant evidence of knowledge and intent, which isn't feasible when the datasets consist of such large collections of art scraped

---

[235] *Id.*

[236] *Id.*

[237] *Id.*

[238] *Id.*

[239] *Andersen v. Stability AI, supra* note 234.

[240] Kevin Madigan, *Top Takeaways from Order in the Andersen v. Stability AI Copyright Case*, Copyright Alliance (Aug. 29, 2024), https://copyrightalliance.org/andersen-v-stability-ai-copyright-case/.

from the internet. Second, based on the nature of generative AI, the identicality requirement under 1202(b) is impossible as "the use of conditioning data to interpolate multiple latent images means the resulting hybrid image will not look exactly like any of the Training Images that have been copied into these latent images,"[241] However, every output from the product is derived exclusively from the training images, and thus, artists' works are still contributing without their permission.[242]

Also noteworthy is the plaintiffs' strategic dismissal of the DMCA claims after submitting their second amended complaint.[243] In Judge Orrick's order in August 2024, while dismissing the plaintiffs' DMCA §1202(b) claims, he also granted leave to amend as long as they filed the second amended complaint without the DMCA-alleging plaintiffs.[244] Given that excluding the claims could potentially compromise the plaintiff's rights to seek reconsideration, in the second amended complaint admitted on October 31, 2024, the plaintiffs initially included the plaintiffs and then voluntarily dismissed with prejudice all DMCA claims in exchange for the defendants agreeing not to challenge future reconsideration of those claims based on their omission from the second amended complaint.[245] The reason for this is because the basis of Judge Orrick's decision to dismiss these claims relies on the outcome of Doe 1 v. GitHub, Inc., which is currently on interlocutory appeal in the Ninth Circuit.[246] If overturned, this strategic move by the plaintiffs will enable reconsideration by the courts of their dismissal of the DMCA claims and enable the court to provide further insight on the issue.[247]

Tags: DMCA, Generative AI, Copyright, 17 U.S.C. Section 1201(a), 17 U.S.C. Section 1201(b), Stability AI, Mid-Journey, DeviantArt, Runaway, DreamUp.

---

[241] *Andersen*, 700 F. Supp. 3d at 861.

[242] *Id.*

[243] James Gatto & Sam Smith, *Andersen Plaintiffs Strategically Dismiss § 1202(b) Claims Pending Interlocutory Appeal in Github Case*, JD SUPRA (Nov. 27, 2024), https://www.jdsupra.com/legalnews/andersen-plaintiffs-strategically-3901814/#:~:text=November%2027%2C%202024-,Andersen%20Plaintiffs%20Strategically%20Dismiss%20%C2%A7%201202(b)%20Claims%20Pending,Interlocutory%20Appeal%20in%20Github%20Case&text=In%20a%20strategic%20move%20to,with%20prejudice%20all%20DMCA%20claims.

[244] *Id.*

[245] *Id.*

[246] Jack Broitman & Stuart Levi, *Recent Decisions on Whether AI Training Violates the Digital Millennium Copyright Act*, JD SUPRA (Dec. 9, 2024), https://www.jdsupra.com/legalnews/recent-decisions-on-whether-ai-training-1219677/.

[247] *Id.*

C.   *Concord Music Group, Inc. V. Anthropic PBC, 2024 US Dist. LEXIS 110598, 1, (M.D.Tenn. 2024).*[248]

The plaintiffs in this case are Concord Music Group, Universal Music Publishing Group, and ABKCO Music Inc, all of whom are leading music labels within the industry.[249] The defendant in this case is Anthropic PBC, an AI safety and research company based in California, who produced the AI assistant "Claude." [250] AI assistants like Claude are used by clients to generate work-product in response to entered prompts. The plaintiffs filed suit against Anthropic PBC for alleged copyright infringement and DMCA § 1202(b) violations for 1) allegedly using copyrighted lyrics to train its generative AI product and 2) copying and distributing those lyrics through Claude's output.[251] The plaintiffs are upset because, Claude was trained using the copyrighted lyrics without permission to use them. The plaintiffs allege that Claude used upwards of 500 song lyrics by major artists including Katy Perry, the Rolling Stones, and Béyonce to train its system. In addition to injunctions, the plaintiffs sought $150,000 per infringement, or $75 million in willful infringement damages.[252] This case was initially filed in October 2023. To train an AI assistant, companies use textual inputs from real-world examples to demonstrate what an acceptable response would be to the AI thinking program. From there, the AI program will generate similar examples based on the instructions given. In this case, the Plaintiffs allege that copyrighted lyrics were used without permission to train Claude. If this were to be true, Anthropic PBC would be in violation of the Plaintiffs' copyright as they utilized lyrics for a commercial purpose (training Claude and reproducing lyric products) without paying for the rights or licensing to use the copyright holders, unless fair use or some other defense proved successful.

Anthropic defense is that it allegedly has implemented internal safeguards within Claude to prevent copyright infringement. There is no public elaboration as to what those safeguards or guardrails entail.   A motion for preliminary injunctions was filed on November 16, 2023 to establish guardrails to prevent Claude's alleged copyright infringement going forward. [253] This preliminary injunction sought for Anthropic to take two actions: one, to cease using copyrighted materials to train its Claude interface; and two, "maintain

---

[248] Hallie Feinman, Tulane Law School, JD expected May 2026.

[249] *Concord Music Group, Inc. v. Anthropic PBC (3:23-cv-01092),* COURT LISTENER, (Jan 24., 2025), https://www.courtlistener.com/docket/67894459/concord-music-group-inc-v-anthropic-pbc/

[250] *See infra* Part II, Subsection C.

[251] Concord Music Grp. v. Anthropic PBC, No. 3:23-cv-01092 (M.D.Tenn. 2024).

[252] Mandy Dalugdug, *Universal Music, Concord Music, ABKCO welcome court-approved 'guardrails' in Anthropic AI dispute – as legal battle continues,* MUSICBUSINESS WORLDWIDE (Jan. 6, 2025), https://www.musicbusinessworldwide.com/universal-music-concord-music-abkco-welcome-court-approved-guardrails-in-anthropic-ai-dispute/.

[253] *Concord Music Grp.*, 738 F.Supp. 3d.

[Anthropic's] already-implemented guardrails."[254] The case was transferred from the Middle District of Tennessee to the Northern District of California on June 24, 2024.

The plaintiffs renewed their motion for preliminary injunction in the Northern District of California in August 2024.[255] Oral argument was heard by Judge Eumi K. Lee on the topic of this preliminary injunction on November 25, 2024. Also in August 2024, Anthropic filed a motion to dismiss the plaintiff's secondary liability claims and DMCA claims as the plaintiffs "fail to allege even one predicate act of infringement by a third party" using Claude's AI interface. Further, Anthropic states that "the only specific examples provided in the Complaint reflect [p]laintiffs' attempts to generate alleged copies of their own lyrics." ding Anthropic's motion to dismiss, although no transcript of the event is available.

Anthropic filed a stipulation and proposed order regarding the preliminary injunction on December 30, 2024. This stipulation was filed after both parties reached agreement regarding Anthropic's maintaining its already-implemented guardrails.[256] Parties also agreed that Plaintiffs may alert Anthropic through writing of any shortcomings or failings of the Claude guardrails, and Anthropic will respond in good faith. On January 2, 2025, Judge Lee entered the stipulation and proposed order.[257] The order does not resolve the remainder of the plaintiffs' motion seeking to enjoin Anthropic for alleged use of infringing copyrighted lyrics to train its Claude systems.[258] Organizations writing on this case include Baker Hostetler, Pillsbury Law, Quinn Emmaneul law firms, as well as CaseText, Court Listener, and a brief article written by Reuters early on in the case's life.[259] Most writing on this case is factual reporting of the case's details without providing commentary.[260] As of January, the case's story has recently been picked up by music law publications such as Music Business Worldwide and The Verge. These music articles are still factual recollections and contain little in regard to legal speculation.

---

[254] *Id.*

[255] *Client Alert: Concord Music Group, Inc. et al. v. Anthropic PBC - Interim Development Narrows Issues in Pending Preliminary Injunction Motion,* QUINN EMANUEL TRIAL LAWYERS (Jan. 7, 2025), https://www.quinnemanuel.com/the-firm/publications/client-alert-concord-music-group-inc-et-al-v-anthropic-pbc-interim-development-narrows-issues-in-pending-preliminary-injunction-motion/.

[256] *Id.*

[257] *Id.*

[258] *Id.*

[259] *Id.*

[260] Blake Brittain, *Music publishers fire back at Anthropic in AI copyright lawsuit,* REUTERS (Feb. 15, 2024), https://www.reuters.com/legal/litigation/music-publishers-fire-back-anthropic-ai-copyright-lawsuit-2024-02-15/.

Tags: music, lyrics, Anthropic, DMCA, 17 U.S.C. Section 1202(b), guardrails, Concord Music Group, Universal Music Publishing Group, ABKCO Music Inc,, copyright infringement, Claude, AI-assistants

D.   Doe v. GitHub, Inc., No. 22-cv-06823-JST (N.D. Cal. May 11, 2023)[261]

Here, the plaintiffs, collectively referred to as "Doe"[262] allege that the defendants, Microsoft, GitHub, and OpenAI, broke open-source software licenses, and violated the DMCA in the creation of Codex and Copilot. This is a class action lawsuit brought by the multitude of plaintiffs, who are all software developers, against Microsoft, GitHub, and OpenAI alleging that these sites used copyrighted materials from the plaintiffs to create two new programs: Codex and Copilot.[263] GitHub is a Microsoft owned platform that allows creators to produce and share code with each other within the platform.[264] By default, that means that GitHub and its parent company Microsoft have access to all of the code created on their systems.

Codex is an OpenAI model that forms the base and powers Copilot, which is GitHub's pair programmer (AI).[265]   GitHub is a collaborative, cloud-based program that allows users to work together to create code as well as store and share it.[266] Copilot is a form of Generative AI programming that allows for a user to write faster, use various Microsoft applications, or specifically for use on GitHub helping users write code faster.[267]

The plaintiffs are a group of programmers who all used GitHub to store their code and collaborate with others. This group of anonymous individuals had 22 claims against GitHub, Copilot, and OpenAI, all owned by Microsoft.[268] The main claims being that GitHub allowed for the Copilot program to train using the software and codes that the plaintiffs had created and stored on their software.[269]

---

[261] Whitney McBay, Tulane Law School, J.D. expected May 2025.

[262] There are five anonymous plaintiffs in this case which is why they are all collectively referred to as "Doe."

[263] *Doe v. GitHub, Inc.*, 22-cv-06823-JST (N.D. Cal. May 11, 2023).

[264] Daniel R. Mello, Jr. et al., *Insights from the Pending Copilot Class Action Lawsuit*, FINNEGAN (Oct. 4, 2023),
https://www.finnegan.com/en/insights/articles/insights-from-the-pending-copilot-class-action-
lawsuit.html#:~:text=GitHub%20Inc.,landscape%20of%20AI%2Dassisted%20coding.

[265] *Id*.

[266] *About GitHub and Git*, GITHUB, https://docs.github.com/en/get-started/start-your-journey/about-github-and-git#about-github (last visited Feb. 5, 2025).

[267] *Cursor vs GitHub Copilot*, BUILDER.IO, https://www.builder.io/blog/cursor-vs-github-copilot (last visited Feb. 5, 2025).

[268] Emma Roth, *The Developers Suing Over GitHub Copilot got Dealt a Major Blow in Court*, THE VERGE (July 9, 2024), https://www.theverge.com/2024/7/9/24195233/github-ai-copyright-coding-lawsuit-microsoft-openai.

[269] *Id*.

The plaintiffs further allege that Copilot does not adhere to the Open-Source Software (OSS) licenses on GitHub that govern plaintiffs' code.[270] OSS is used to manage the software and code of many software developing systems, like GitHub and Copilot.[271] Plaintiffs made 3 primarily alleging claims in the OSS license issue. These state that Copilot did not or failed to provide: 1) acknowledgement and attribution to the owner of the code, 2) notice of copyright, and 3) the explicit terms of the license which did not include the express OSS licensing that allows for Copilot to create derivative works based on the information input into its system by creators.[272]

This case is unique in that while it displays itself as a copyright case against an LLM provider, it does not actually include any claims of indirect or direct copyright infringement.[273] Among the 22 claims, some of which dabble in indirect copyright infringement, the majority were thrown out. The main amended claims only involve ▢DMCA violations, breach of contract, and multiple torts.▢ The California judge presiding over the case only left two primary claims standing among the 22 filed, a breach of contract violation, and an open-source license violation.[274] Most of the claims were dismissed with prejudice, so the creators cannot refile any of those claims.

The case was on hold pending the plaintiff's interlocutory appeal of the dismissal of its DMCA §1202 claims but has since been decided[275]▢ The Court had agreed with some claims but dismissed others. The plaintiffs believed that they had a valid DMCA claim and that it should not have been dismissed.[276] The key issue is whether Section 1202(b) has an identicality requirement. This requirement means that the current product is identical to the original product except for certain personalized information (name or terms and conditions).[277]

---

[270] *Doe v. GitHub, Inc.*, 22-cv-06823-JST (N.D. Cal. May 11, 2023).

[271] *Id.*

[272] Mello et. al., *Insights from the Pending Copilot Class Action Lawsuit*, FINNEGAN (Oct. 4, 2023),
https://www.finnegan.com/en/insights/articles/insights-from-the-pending-copilot-class-action-
lawsuit.html#:~:text=GitHub%20Inc.,landscape%20of%20AI%2Dassisted%20coding.

[273] *Doe v. GitHub, Inc.*, BAKERHOSTETLER, https://www.bakerlaw.com/the-copilot-litigation/ (last visited Feb. 5, 2025).

[274] *Id.*

[275] *Id.*

[276] *Id.*

[277] Maria Crusey, *Copyright Management Information, 1202(b), and AI, Authors Alliance*, AUTHORS ALLIANCE ,
https://www.authorsalliance.org/2024/10/30/copyright-management-information-1202b-and-
ai/#:~:text=%E2%86%92%20The%20%E2%80%9CIdenticality%E2%80%9D%20Requirement,of%20Nevada%20in%20Oracle%20v. (last visited Feb. 5, 2025).

The plaintiffs claim that this section of the DMCA [278]requirement was relevant for their claims, but the Judge did not agree. He claimed that the code that GitHub had copied from the creators was similar, but not similar enough to the original work of the developers.[279] In a study done of AI and code-generating programs, the Judge quoted a section that said Copilot rarely regurgitates memorized code.[280] This killed the claim that Copilot violated the DMCA by not giving proper credit and attribution to the creators of the code it was suggesting.[281]

The same Court denied the defendants' motions regarding plaintiffs' claims for breach of contract and for violations [282] The plaintiffs claimed that the Copilot application, operated by GitHub, was operating using the stolen code from the plaintiffs but did not include the identifying information.[283] This would violate DMCA 1202(b)(1) and (b)(3) if GitHub and Copilot had altered their CMI in an active way.[284] CMI stands for "copyright management information" which can include things like the author, the title, terms and conditions for work, copyright owner, or just any identifying information that would be set forth in a copyright notice.[285] A CMI can often be found before a block of code typically. This would be a violation of DMCA 1202(b)(1) and (b)(3) if the plaintiffs can prove that the defendants' program was specifically designed to actively alter or remove any CMI that was affixed to the code coming through its systems.[286]

The defendants tried to file a motion to dismiss because they alleged that they did not plead sufficiently regarding the defendant's role in their misconduct[287] However, the plaintiffs were able to point out that Copilot and Codex were specifically designed to remove any CMI from licensed code that they reproduced and they were taught to immediately remove CMI (or ignore it) when reproducing for training.[288] And so, these two claims continue as the Court found that the plaintiffs had pled sufficiently to support this claim. The Court instead took into account the plaintiffs leave to amend and dismissed the remaining[289] Their claims for civil conspiracy and declaratory relief were also rejected because they did not meet the requirements. [290]

Regarding unjust enrichment and unauthorized use, the Court noted that unjust enrichment is not automatically preempted by copyright and a state law tort claiming unauthorized use of a product or system is also not automatically

---

[278] *Id.*

[279] Roth, *supra* note 268.

[280] *Id.*

[281] *Id.*

[282] *Doe v. GitHub, Inc.*, 22-cv-06823-JST (N.D. Cal. May 11, 2023).

[283] Crusey, *supra* note 277.

[284] *Id.*

[285] *Id.*

[286] *Id.*

[287] *Doe v. GitHub, Inc.*, 22-cv-06823-JST (N.D. Cal. May 11, 2023).

[288] *Id.*

[289] *Id.*

[290] *Id.*

preempted by copyright[291] The defendant's claim that the plaintiffs state law claims are preempted by Section 301 of the Copyright Act and note that most of these were dismissed, except for unjust enrichment.[292] The plaintiffs don't address their arguments for preemption, but instead state that their state law claims are very different from their Copyright Act claims because they are about both the unauthorized use and production of the code.[293] On their preparation of derivative works, the Court found that their unjust enrichment claim was properly preempted by the Copyright Act. The claim was still dismissed with leave to amend. The opening brief for the plaintiffs is due in March 2025.[294] This case has become a training ground for the future of AI cases in courts.[295] The ruling also hammers home that plaintiffs will have to produce some form of compelling evidence and actual examples of product reproductions in order to have standing to sue in the future.[296]

Tags: GitHub, AI training, code, open-source software (OSS), Digital Millennium Copyright Act (DMCA), 17 U.S.C Section 1202(b)1); 17 U.S.C. Section 1202(3), Copilot, unjust enrichment, unauthorized use of a product or system (under state law) copyright preemption, Codex.

E.   *Getty Images (US) v. Stability AI., Inc., No. 1:23-cv-00135-JLH (D. Del. Feb. 3, 2023)*[297]

    This case was brought by the plaintiff, Getty Images, Inc. (Getty Images), against defendant Stability AI.[298] Getty Images Inc. is a visual media company that is commonly known for their large digital catalog of stock images.[299] Stability AI is an open-source generative AI company that develops and trains AI models for different applications including audio, language, code, and imaging.[300] Stable Diffusions is the advanced image generative AI model that is used on the platform, DreamStudio to create images using text descriptions.[301]
    Getty Images accuses Stability AI of infringing on more than 12 million photographs, their metadata and associated captions in training and building

---

[291] Mello et. al., *Insights from the Pending Copilot Class Action Lawsuit*, Finnegan (Oct. 4, 2023)

[292] *Id*.

[293] *Id*.

[294] BakerHostetler, *supra* note 273.

[295] *Id*.

[296] *Id*.

[297] Jamika P. Mack, Tulane University Law School-Class of 2025

[298] Complaint, *Getty Images v. Stability AI., Inc.*, 1:23-cv-00135-JLH, at *1 (D. Del, Feb. 3, 2023).

[299] *Id.* at 1-2.

[300] *Id.* at 3.

[301] *Id.*

Stable Diffusion and Dream Studio.[302] Getty Images further alleges that because Stable Diffusion used Getty Images to train their AI models; the Getty Images watermarks are still evident on the AI output.[303] Getty Images recognizes AI's ability to stimulate creative endeavors.[304] Therefore, Getty Images provided licenses to technology innovators for the purposes of training AI systems in a manner that respects personal and intellectual property rights.[305] According to Getty, Stability AI did not seek to obtain a license from Getty Images and instead chose to ignore the viable licensing options and long- standing IP legal protections to pursue commercial interests.[306]

There are eight causes of action that Getty Images brought against Stability AI.[307] This includes direct copyright infringement, alleging that Stability AI infringed and will continue to infringe by reproducing Getty Images' copyrighted works and creating derivative works.[308] The plaintiffs also cite DMCA 1202(a) violations for providing false Copyright Management Information by "applying a modified version of Getty Images' watermarks to output generated through use of Stable Diffusion and the Dream Studio interface"[309] They also include DMCA 1202(b) violations for removing Getty Images' watermarks and metadata [310] Finally, the causes of action include trademark infringement, unfair competition, trademark dilution under both federal and Delaware trademark law and violation of Delaware's Uniform Deceptive Trade Practices Act for unauthorized use of Getty Images' trademarks in connection with synthetic images generated by Stable Diffusion and DreamStudio.[311]

The latest judgement occurred on January 15, 2025.[312] The court ruled against the Sixth Claimant's attempt to bring a representative action on behalf of over 50,000 copyright owners, citing the lack of a definitive class definition and the necessity of individualized liability assessments.[313] The court ruled that Stability AI's limited admissions regarding the use of Getty Images' works made it impossible to determine a specific group of eligible claimants.[314] The court ruled

---

[302] *Id.* at 1.

[303] *Id.* at 4.

[304] *Id.* at 3.

[305] *Id.*

[306] *Id.*

[307] *Getty Images v. Stability*, BAKERHOSTETLER, https://www.bakerlaw.com/getty-images-v-stability-ai/ (last visited Feb. 11, 2025).

[308] *Id.*

[309] *Id.*

[310] *Id.*

[311] *Id.*

[312] Racheal Muldoon et. al, *Getty v Stability AI: A 'tantalising glance' of what's to come for AI firms and creators*, Charles Russell Speechlys (Jan. 16, 2025), https://www.charlesrussellspeechlys.com/en/insights/expert-insights/financial-services/2025/getty-v-stability-ai-a-tantalising-glance-of-whats-to-come-for-ai-firms-and-creators/

[313] *Id.*

[314] *Id.*

that the claimants' argument—that the defendant possessed the necessary
knowledge to identify infringing works—was insufficient to justify a
representative action. [315] The court ruled against the claimant's informal
application under CPR 19.3, determining that it lacked sufficient evidence and
would create an excessive litigation burden.[316] The court ruled that mass copyright
claims in AI training cases require structured legal mechanisms rather than broad
representative claims.[317] The court ruled that the parties must now resolve case
management issues in preparation for the first trial in June 2025.[318]

Shortly after this lawsuit, Getty partnered with NVIDIA to develop its own
AI-photo generation tool called Generative AI by Getty Images. [319] Getty's
motivation behind entering this new market is to give creatives and copyright
owners peace of mind that any images generated and licensed through their new
platform would be backed by their "uncapped legal indemnification".[320] Getty
alleges that its new tool is unlike other AI-photo generation tools because the legal
rights have been cleared prior to training the models. [321] Further, Getty is
guaranteeing that users would have worldwide, perpetual, and non-exclusive use
of the images they create and that those AI-generated creations would not be
available for others to use.[322]

Commentators are skeptical of Getty's new tool.[323] There are many concerns
about photographers not being properly compensated for their work, ethical issues
regarding the use of copyrighted material to train AI models in general, and the
poor quality of the newly-generated content. [324] Getty says that contributing
photographers are paid for any of their work that was used to train the model, with
compensation being about 30 cent to a dollar for every dollar generated.[325] Getty
claims that this is made possible from the subscription fee that Getty charges to
customers for using the tool.[326] Skeptics are worried that fair compensation is an
issue because it is not clear which photos in the training set would be given credit

---

[315] *Id.*

[316] *Id.*

[317] *Id.*

[318] *Id.*

[319] *Generative AI by Getty Images*, GETTY IMAGES, https://developers.gettyimages.com/ai-generation/ (last visited Feb. 11, 2025).

[320] *Id.*

[321] Emilia David, *Getty made an AI generator that only trained on its licensed images*, THE VERGE (Sep. 23, 2023), https://www.theverge.com/2023/9/25/23884679/getty-ai-generative-image-platform-launch

[322] *Id.*

[323] Lauren Goode, *Getty Images Plunges Into the Generative AI Pool*, WIRED (Sept. 25, 2023), https://www.wired.com/story/getty-images-generative-ai-photo-tool/.

[324] *Id.*

[325] David, *supra* note 321.

[326] *Id.*

for generated content.[327] Getty admits that they don't know how to identify what contributor that each generated content derives from but that they utilize technology to determine that.[328] Additionally, some users argue that Getty's AI tool is outperformed by other image generators, such as DALL-E.[329] After entering the same prompt in both Getty's tool and DALL-E and Clipdrop, the user described the results as being "noticeably less interesting aesthetically and from a narrative perspective, and they were overall rather obvious to the point of being bland."[330] Overall, the consensus seems to suggest that Getty would need to continually refine its model to ensure high quality output and ensure that contributing creatives are being legally protected and compensated for their work.[331]

Tags: Getty Images, Stability AI, AI model training, Digital Millennium Copyright Act, stock photos, Generative AI by Getty Images, copyright material, AI-image generator

F. *In re Google Generative AI Copyright Litig., No. 5:23-CV-03440-EKL (N.D. Cal. Dec. 20, 2024) (Consolidated Class Action Complaint).*[332]

The Plaintiffs in the current consolidated class action are a group of authors and visual artists, including New York Times best-selling authors like Steve Almond and cartoonists like Sarah Andersen, who are suing for direct and vicarious copyright infringement. Plaintiffs allege that Google infringed their copyrights by copying their works without authorization to train its generative AI models, Gemini and Imagen, thereby harming their licensing revenue and market value.

In Re Google is a consolidated action of two putative class action suits; *Leovy v. Google,* and *Zhang v. Google.*[333] In July 2023, the eight original Plaintiffs in the *Leovy* action had chosen to remain anonymous. Plaintiffs included a New York Times best-selling author and investigative journalist, a six-year- old and thirteen- year-old minor, and other Google and Gmail users filed a class action claiming that "Google has been secretly stealing everything ever created and shared on the internet by hundreds of millions of Americans,"[334] and that specifically, Google directly infringed the rights of copyright owners when it

---

[327] Goode, *supra* note 323.

[328] *Id.*

[329] Tiernan Ray, *I tried Getty's new AI image generator, and it doesn't compare to DALL-E*, ZDNET (Jan. 9, 2024), https://www.zdnet.com/article/i-tried-gettys-new-ai-image-generator-and-it-doesnt-really-compare-to-dall-e/.

[330] *Id.*

[331] *Id.*

[332] Isa Reyes-Klein, J.D.,Tulane University Law School, expected graduation May 2026

[333] Complaint, *In re Google Generative AI Copyright Litigation*, 5:23-cv-03440, (N.D. Cal. Jul 11, 2023) ECF No. 1.

[334] *Id.*

collected and used the works to train its AI technology.[335] The original complaint names Google DeepMind--an Alphabet subsidiary AI research laboratory, [336] Google LLC ,[337] and Google's parent company Alphabet[338] as defendants, alleging that Google secretly collected and used personal, creative, and copyrighted data from millions of internet users without consent to develop its AI products, including Bard which is now known as the AI chatbot assistant Gemini which competes with OpenAI's ChatGPT.[339] The ten alleged causes of action include unfair competition, invasion of privacy, intrusion upon seclusion, larceny/receipt of stolen property, conversion and unjust enrichment based on Google's alleged use of the Plaintiffs' private and personal information that was collected or "scraped" without the consent of the users or their knowledge.[340] Plaintiffs further assert that they were not made aware that by using Google products their information could be collected without their consent.[341] In the original action, Plaintiffs sought an injunction to stop the data collection, allow users to opt out, and demand deletion or compensation for misused data.[342]

The Plaintiffs further alleged violation of the Digital Millennium Copyright Act ("DMCA") for distributing derivative works without copyright management information ("CMI") in their copyrighted materials.[343] Plaintiffs alleged such direct infringement arose from the scraping and use of copyrighted works to train Google's AI products.[344] Plaintiffs cite a study which found that in the data sets which were used to train Gemini, the copyright symbol appeared more than 200 million times, suggesting a vast number of works used to create and train the chatbot were protected by copyright.[345] Namely, Plaintiffs alleged that Google

---

[335] *Id.*

[336] *Announcing Google DeepMind*, Google DeepMind (Apr. 20, 2023), https://www.deepmind.com/blog/announcing-google-deepmind (The generative AI research and development company describes artificial intelligence technologies as having the potential to "be one of humanity's most useful inventions.").

[337] *Google LLC*, Bloomberg, https://www.bloomberg.com/profile/company/8888000D:US#xj4y7vzkg (last visited July 10, 2023) (Google LLC was restructured as a subsidiary of Alphabet Inc. in 2015.).

[338] Alphabet Inc. is a multinational technology company that owns Google and other companies since 2015 when Google reorganized. *See Deepfakes: The Synthetic Videos Fooling Us All*, BBC News (Nov. 27, 2019), https://www.bbc.com/news/technology-50656803.

[339] *Gemini Apps FAQ*, Gemini, https://gemini.google.com/faq (last visited Feb. 11, 2025).

[340] *Id.*

[341] *Id.*

[342] Complaint, 54, *In re Google Generative AI Copyright Litigation*, 5:23-cv-03440, (N.D. Cal. Jul 11, 2023) ECF No. 1.

[343] *Id.*

[344] *Id*

[345] Kevin Schaul et al., *Inside the Secret List of Websites that Make AI like ChatGPT Sound Smart*, Wash. Post (Apr. 19, 2023), https://www.washingtonpost.com/technology/interactive/2023/aichatbot-learning/.

reproduced Plaintiff's copyrighted materials when they were used to train Gemini, a direct competitor to Open AI's "Chat GPT" and Google's Imagen an AI photo editing assistant[346] Plaintiffs further assert Defendants scraped at least 1.56 trillion words of "public dialog data and other public web documents" from websites like LAION and Common Crawl that provide free data for research specific purposes[347] rather than purchasing internet user data sold like any other content or property.[348] On September 19, 2023, Plaintiffs voluntarily dismissed Alphabet Inc. and Google DeepMind.[349] Google LLC filed motions to dismiss for failure to state claims to which Plaintiffs filed an amended complaint. Subsequently, on May 28, 2024, Google notified the court of a recent decision in *Cousart v. OpenAI LP*, where the court dismissed a complaint for violating Federal Rule of Civil Procedure 8 by being excessively long and containing irrelevant information—an argument Google makes in its own motion to dismiss in *Leovy,* further noting that six of the *Cousart* plaintiffs are represented by the same counsel as the *Leovy* plaintiffs.[350]

Three days later, Google LLC filed a motion to relate *Leovy* and *Zhang et al. v. Google* LLC, to which Plaintiffs repeatedly opposed.[351] However, on July 23, 2024, the court granted Google's motion to relate the two cases and permitted

---

[346] *Plagiarism, AI & ChatGPT: ChatGPT: How does work?*, UNIV. LIBR., https://guides.library.txstate.edu/c.php?g=1321038&p=9718369#:~:text=What%20exactl y%20is%20chat%20GPT,is%20similar%20to%20human%20conversation (last visited Feb. 11, 2025) (explaining that a chatbot is a type of artificial intelligence that can understand and generate natural language text." Chat bots are trained "on large amounts of text data" and use an algorithm "that is designed to mimic the way the human brain works." The chatbot is trained on a large dataset of text, such as articles or conversations." It then uses this data to learn the patterns and structure of language. Once it has learned enough about language, it can generate its own text based on a given prompt or topic. "For example, if you ask chatGPT about the weather. It might generate a response like "The weather today is sunny and warm, with a high of 75 degrees." It is able to generate this response because it has learned about weather patterns and how to describe them in natural language.").

[347] Such sites serve as repositories for over 250 billion web pages for over 18 years. S*ee* Romain Beaumont, LAION-5B: A New Era of Open Large-Scale Multi-Model Datasets, LAION (Mar. 31, 2022), https://laion.ai/blog/laion-5b/; *See also* Homepage, COMMON CRAWL, https://commoncrawl.org/ (last visited Feb. 11, 2025).

[348] Calvin Wankhede, *What Is Google's Bard AI? Here's Everything You Need to Know,* ANDROID AUTHORITY (Mar. 22, 2023), www.androidauthority.com/google-bard-chatbot-3295464/ (explaining that 'bots,' or robot applications scan and copy the information on web pages then store and index what is described as an unimaginable extent of personal information.).

[349] Notice of Voluntary Dismissal, J.L. v. Alphabet Inc., No. 3:23-cv-03440-AMO (N.D. Cal. Sept. 19, 2023).

[350] Cousart v. OpenAI LP, No. 3:23-cv-04557-VC (N.D. Cal. May 24, 2024).

[351] Consolidated Class Action Complaint, In re Google Generative AI Copyright Litig., No. 5:23-CV-03440-EKL (N.D. Cal. Dec. 20, 2024).

Plaintiffs to Respond to Google's Renewed Motion to Relate.[352] The *Zhang* Plaintiffs opposed Google's motion to relate their copyright infringement case which concerned visual artists' work used in image generation AI with the *Leovy* case which concerned authors' work used in language model AI, arguing the cases involve different parties, different types of copyrighted work, and different AI models, thus lacking substantial similarity.[353] However, on August 5, 2024, the court ordered the *Zhang* and *Leovy* cases be related which was followed by the Court granting Google's motion for Joint Stipulation to Consolidate *Leovy* and *Zhang* on Oct. 28, 2024, because both cases involved similar claims against Google regarding their use of copyrighted material.[354]

   At present, *In re Google Generative AI Copyright Litigation*, is a consolidated class action complaint with eleven named Plaintiffs who allege that Google copied millions of copyrighted works in its Generative AI products like Gemini and permanently embedded copies of these works within these models.[355] Plaintiffs include New York Times Best-Selling author Steve Almond, who owns nine registered copyrights for literary works and alleges that at least six of which "Google took, used, copied, and/or reproduced without license or authorization."[356] As a result, Almond alleged he suffered direct market harm from the revenue he could have received if Google had properly licensed his works for AI training and market value diminution due to Google's "mass appropriation of literary works" which he alleges has depressed the overall market for fiction and nonfiction writers.[357] Similarly other Plaintiffs include New York Times best-selling authors, illustrators, cartoonists, photographers, visual artists, who are all registered owners of one or more copyrighted works.[358] The plaintiffs all alleged that Google's unauthorized copying and mass appropriation resulted in direct market harm and market value diminution in their respective

---

[352] Zhang Plaintiffs' Opposition to Google's Renewed Administrative Motion to Consider Whether Cases Should be Related, Leovy v. Google LLC, No. 3:23-cv-3440-AMO (N.D. Cal. July 30, 2024).

[353] Zhang Plaintiffs' Opposition to Google's Renewed Administrative Motion to Consider Whether Cases Should be Related, Leovy v. Google LLC, No. 3:23-cv-3440-AMO (N.D. Cal. July 30, 2024).

[354] At first, the court denied Google's motion to relate the original case *Leovy v. Google* with *Zhang v. Google*, however, after the privacy and state law claims were dropped by both Plaintiffs, Google stipulated that both the putative class actions focused on copyright infringement claims for training its AI models, including Gemini and Imagen, and that Google would anticipated claiming the same defense in both cases. Joint Stipulation and Proposed Order Consolidating Cases, Leovy v. Google LLC, No. 5:23-cv-03440-EKL, & Zhang v. Google LLC, No. 5:24-cv-02531-EKL (N.D. Cal. Oct. 28, 2024).

[355] Consolidated Class Action Complaint, In re Google Generative AI Copyright Litig., No. 5:23-cv-03440-EKL, 1 (N.D. Cal. Dec. 20, 2024) ().

[356] *Id.* at 6.

[357] *Id.*

[358] *Id*. at 6-20.

industries.[359] The Plaintiffs determined their copyrighted works had been used and copied without proper license by listing the websites their copyrighted works were found on that were alleged to have been used to train Google's AI models. The consolidated class action complaint claims that Google's embedding of the Plaintiffs works into the Generative AI Models, Google has irreversibly entangled the Plaintiff Works with its commercial products, which in effect has stripped Plaintiffs of their exclusive rights to control the copying and distribution of their registered works under the Copyright Act.[360]

Presently, Plaintiffs are waiting for defendants to respond to the consolidated class action complaint. Websites such as those hosted by Baker & Hostetler LLP, Law 360, and Ropes and Gray are tracking the AI litigation across the US in an objective and concise manner.

Tags: Google, Gemini, AI Training, OpenAI, Copyright Management Information, ChatGPT

G.   Huckabee v. Bloomberg, No. 1:23-cv-09152 (S.D.N.Y. October 17, 2023).[361]

Former Arkansas governor Mike Huckabee and others are suing Bloomberg over its alleged use of a dataset containing unlawfully obtained works by the authors to train its AI model. The plaintiffs claim Bloomberg knowingly engaged in direct copyright infringement of their works and have cited specific works by each author present in the dataset, as well as reports from Bloomberg acknowledging that the dataset was used for training an early iteration of its AI model. Bloomberg claims fair use.

Plaintiff-authors Mike Huckabee, Relevate Group, David Kinnaman, Tsh Oxeneider, Lyla Terkeurst, and John Blasé are suing Bloomberg L.P. and Bloomberg Finance L.P. (hereinafter "Bloomberg")[362], alleging Bloomberg engaged in direct copyright infringement when it trained its Large Language Model (LLM), BloombergGPT, on unlicensed copies of the plaintiffs' copyright-protected works using the Books3 dataset.[363]

---

[359] *Id.*

[360] *Id.*

[361] Maddy Marlowe, Tulane Law School, expected May 2025.

[362] Initially, this lawsuit involved defendants Meta Platforms, Inc., Microsoft Corporation, and The EleutherAI Institute; however, Meta and Microsoft were severed as defendants by the court and moved to California while this Bloomberg suit remains in the S.D.N.Y.. The EleutherAI Institute was voluntarily dismissed as a defendant. The present lawsuit is now limited to the listed Bloomberg entities. Class Action Complaint, 1, *Mike Huckabee, Relevate Group, David Kinnaman, Tsh Oxeneider, Lyla Terkeurst, and John Blasé v. Meta Platforms, Inc., Bloomberg L.P., Bloomberg Finance L.P., Microsoft Corporation, and The EleutherAI Institute*, No. 1:23-cv-09152 (S.D.N.Y. October 17, 2023).

[363] Plaintiffs' initial complaint also alleged vicarious infringement and DCMA claims, but these were not included in the amended complaint now at issue against Bloomberg. *Status Updates,* Huckabee v. Bloomberg, BAKERHOSTETLER (last accessed 2/12/25), https://www.bakerlaw.com/huckabee-v-bloomberg/.

The Books3 dataset was introduced in 2020 by EluetherAI as part of its larger dataset "The Pile" and comprises information from 183,000 pirated ebooks obtained through data scraping.[364] This dataset became a popular resource for major companies, including Bloomberg, in training their developmental AI models to respond to prompts with natural language. Plaintiffs now allege the illicitly obtained dataset was used by Bloomberg to train its LLM, BloombergGPT, despite Bloomberg's "full knowledge and understanding" that the dataset contained unlicensed copyrighted works, such as those by the Plaintiff-authors.[365] BloombergGPT is a generative AI system for financial analysis developed by the company.[366] In its letter of intent to move to dismiss all causes of action, Bloomberg argued the fair use defense. Bloomberg claims the alleged training was part of a not-for-commercial-use research project to develop an LLM for improved performance in financial analysis.[367]

Bloomberg maintains that BloombergGPT is an internal company application and that any alleged use of the Plaintiff-authors' works was limited, private, and conducted by a news-reporting agency for non-commercial research purposes, falling well within the "purpose and character" of fair use.[368] Bloomberg also states the plaintiffs have failed to provide specific evidence of how it infringed or which books were misused for BloombergGPT.[369]

In their initial complaint, plaintiffs cited an announcement by the company acknowledging Books3 was used in training its initial model of BloombergGPT, but that the dataset would not be included when training future models of the LLM.[370] In response to this claim, the plaintiffs argue the use of Books3 to train future models is inconsequential because the nature of LLM development involves building on previous models; therefore, the plaintiffs allege their protected works from Books3 are already baked into even those future versions of BloombergGPT.[371] The plaintiffs maintain Bloomberg has already and will continue to benefit from their unlicensed, copyright-protected works.

In its motion to dismiss, Bloomberg alleged the plaintiffs' claims are insufficient and even if accepted as true, are rebutted by plaintiffs' own amended

---

[364] *Id.*

[365] *Id.*

[366] Blake Brittain, *Bloomberg asks US court to toss copyright lawsuit over AI training*, REUTERS.COM (last updated March 25, 2024), https://www.reuters.com/legal/litigation/bloomberg-asks-us-court-toss-copyright-lawsuit-over-ai-training-2024-03-25/.

[367] *Status Updates,* Huckabee v. Bloomberg, BAKERHOSTETLER, https://www.bakerlaw.com/huckabee-v-bloomberg/ (last visited Feb. 12, 2025).

[368] Brittain, *supra* note 366.

[369] *Id.*

[370] Class Action Complaint, 18, Mike Huckabee, et.al. v. Meta Platforms, Inc., Bloomberg L.P., Bloomberg Finance L.P., Microsoft Corporation, and The EleutherAI Institute, No. 1:23-cv-09152 (S.D.N.Y. October 17, 2023).

[371] *Id.*

complaint recognizing Bloomberg as a "provider of financial news and analysis", arguing this acknowledgement clearly demonstrates the BloombergGPT project was a fair use.[372]

The lawsuit is currently pending. Following Bloomberg's motion to dismiss, plaintiffs responded with arguments that the motion failed to deny plaintiffs' ownership of copyright in those works specifically identified in the amended complaint as part of the Books3 dataset used by Bloomberg, and that Bloomberg's fair use defense is not appropriate at this stage because it requires facts not included in the complaint and assumptions drawn in Bloomberg's favor.[373] The most recent filing in this case was Bloomberg's response to the plaintiffs' response to its motion to dismiss, wherein Bloomberg argued once again failure to allege claims with sufficient, specific facts indicating a basis for relief. Bloomberg stated in its response that the plaintiffs are engaged in a "mere fishing expedition without a substantiated legal claim."[374]

As it stands, the lawsuit is still pending in the Southern District of New York, and there has been no ruling on Bloomberg's motion to dismiss. Notably, there is a lack of secondary sources providing opinion on the lawsuit or the arguments raised by the parties. However, the fair use defense raised by Bloomberg is reflective of a theme in much AI litigation over copyright infringement. The suit is being monitored and reported by Baker Hostetler and Reuters through status updates on AI litigation around the country.

Tags: Bloomberg, BloombergGPT, Mike Huckabee, Books3, copyright infringement, fair use

H. The Intercept Media and Raw Story Media v. OpenAI, No. 1:24-cv-01515, (S.D.N.Y. Feb. 28, 2024) [375]

These are three news organizations alleging DMCA violations in training ChatGPT, specifically involving the removal of Copyright Management Information. On February 28, 2024, Intercept Media, Raw Story media, and AlterNet, three news media corporations, filed suit against both OpenAI and Microsoft. [376] with claims the defendants are adversarial to journalism organizations. Raw Story Media claims it is the largest independent progressive political news website.[377] AlterNet is a publisher with a focus on civil rights,

---

[372] *Status Updates, supra* note 367.

[373] *Id.* Plaintiffs also maintain that all allegations must be drawn in their favor at the Motion to Dismiss stage, and cited to precedent on fair use defense at this stage to argue they should be allowed to proceed to discovery to challenge Bloomberg's defense. *Id.*

[374] *Id.*

[375] Layla Ghiai, Tulane Law School, JD, expected graduation, 2025.

[376] Complaint at 1, The Intercept Media, Inc. v. OpenAI, Inc., No. 1:24-cv-01515, (S.D.N.Y. 2024); Complaint at 1, Raw Story Media, Inc. v. OpenAI Inc., No. 1:24-cv-01514, (S.D.N.Y. 2024)

[377] *Raw Story Media, Inc.*, No. 1:24-cv-01514 at 2-3

social justice, culture, health, and the environment. Defendants OpenAI and Microsoft jointly are the creators and trainers of ChatGPT AI products.[378]

The claims that the plaintiffs make against OpenAI and Microsoft are violations of the integrity of Copyright Management Information statutes within the DMCA (Digital Millennium Copyright Act), which is part of the 1976 Copyright Act. In particular, plaintiffs argue that OpenAI and Microsoft removed the copyright management information (CMI) of various articles in order to train ChatGPT. This removal of CMI information then facilitated and enabled users of ChatGPT to enact copyright infringement without their own knowledge.[379] None of the media corporations claimed OpenAI committed copyright infringement itself. In addition to damages, both Intercept and Raw Story sought an injunction requiring OpenAI and Microsoft to remove all copies of its copyrighted works that do not maintain the original CMI.

On April 15, 2024, Microsoft and OpenAI filed a motion to dismiss.[380] In the motion, Microsoft pled that the news media corporations failed to provide examples of copyrighted work that were produced in any ChatGPT outputs despite their claim that the works are being used in the training datasets. Moreover, the defendants argue that there was no evidence of a concrete injury suffered by any of the plaintiffs, therefore not giving the plaintiff's any legal standing. Lastly, Microsoft and OpenAI claimed that plaintiffs failed to overcome the "double scienter" requirement needed to satisfy any DMCA claims. In order to overcome the "double scienter" requirement, a plaintiff must allege that the defendant had removed the CMI information intentionally and that the defendant must have reasonably understood that removal of the CMI may facilitate copyright infringement. Ultimately, the Southern District of New York granted Microsoft and OpenAI's motion for dismissal against Raw Story Media under the grounds that plaintiffs did not have proper standing due to no concrete injury being present. Meanwhile, all but one of Intercept's claims has been dismissed.

Currently, both suits are still ongoing. Raw Story Media Inc. and Alternet have filed for a motion for leave to amend their complaint while OpenAI has answered Intercept's one remaining claim. Although ultimately, these media outlets failed under a standing issue, many experts fear it will be impossible to bring suit against any AI model training especially since proving the concrete injury is difficult.[381] OpenAI currently has partnerships with 38 different

---

[378] *See infra* Part II, Section A.

[379] Complaint at 13, The Intercept Media, Inc. v. OpenAI, Inc., 1:24-cv-01515, (S.D.N.Y.)

[380] *The Intercept Media and Raw Story Media v., OpenAi,* BAKERHOSTETLER, https://www.bakerlaw.com/the-intercept-media-and-raw-story-media-v-openai/ (Last visited Feb 2025).

[381] Kate Knibbs, *OpenAI Scored a Legal Win Over Progressive Publishers—but the Fight's Not Finished,* WIRED, https://www.wired.com/story/opena-alternet-raw-story-copyright-lawsuit-dmca-standing/ (Nov 8, 2024)

corporations, several being news media outlets: TIME, Vox, and The Atlantic to name a few.[382]

Tags: Newspaper, Media, OpenAI, AI Training, Copyright Management Information, ChatGPT

I.   *Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC, 2023 WL 8039640, (N.D. Cal. Nov. 20, 2023).*[383]

Plaintiff authors Sarah Silverman, Richard Kadrey, and Christopher Golden brought a putative class against Meta, a technology company, after it allegedly trained its artificial intelligence software Large Language Model Meta AI (LLaMA) using the plaintiffs' books as training sources without their consent, credit, or compensation.[384] The plaintiffs initially alleged six causes of action, including direct copyright infringement, vicarious copyright infringement, violation of the Digital Millennium Copyright Act (DMCA), unfair competition, unjust enrichment, and negligence.[385] In November 2023, the court dismissed all claims other than the plaintiff's assertion that Meta's alleged unauthorized copying of their books to train LLaMA constituted direct copyright infringement.[386]

The court dismissed the vicarious copyright infringement claim because the plaintiffs did not argue sufficiently that the models created an infringing output, a necessary element for this type of infringement.[387] According to some intellectual property lawyers, plaintiffs' claim for vicarious copyright infringement failed because the complaint did not allege that LLaMA created an output containing protectable expression that recast, transformed, or adapted the books.[388] Since the court found no infringing output, there could not be vicarious infringement.[389] The court dismissed the plaintiffs DMCA claim because it found they did not properly allege that LLaMA distributed the plaintiffs' books outside of their alleged use.[390] The claim for unfair competition was also dismissed by the court because it was preempted by the plaintiffs' direct copyright infringement claim.[391] Lastly, in the opinion of the court, the claims of unjust enrichment and negligence were

---

[382] Jonathan Gillham, *Open AI Partnership List,* ORIGINALITY.AI, https://originality.ai/blog/openai-partnerships (Jan 21, 2025).

[383] Lily Argyle**,** Tulane University Law School, JD expected May 2025.

[384] *Kadrey v. Meta Case Tracker,* BAKERHOSTETLER, https://www.bakerlaw.com/kadrey-v-meta/ (last visited Feb. 5, 2024).

[385] Melanie J. Howard & Keane Barger, *IP/Entertainment Case Law Updates: Kadrey V. Meta Platforms, Inc.,* LOEB & LOEB IP/ENTERTAINMENT CASE LAW UPDATES (Feb. 5, 2025), https://www.loeb.com/en/insights/publications/2023/12/richard-kadrey-v-meta-platforms-inc.

[386] *Kadrey v. Meta Case Tracker, supra* note 384.

[387] LOEB & LOEB IP/ENTERTAINMENT CASE LAW UPDATES, *supra* note 385.

[388] *Kadrey v. Meta Case Tracker, supra* note 384.

[389] LOEB & LOEB IP/ENTERTAINMENT CASE LAW UPDATES, *supra* note 342.

[390] *Kadrey v. Meta Case Tracker, supra* note 385.

[391] *Id.*

preempted by the Copyright Act.[392] In defense of the remaining direct copyright infringement claim, Meta maintains that the use of the plaintiffs' works for training falls under fair use.[393] Based on the court's dismissal of these claims and Meta's defense, it seems that the current and potential future plaintiffs are more likely to be successful in cases where the copyright material used to train AI tools is substantially similar to the outputs of the AI tools.[394]

In July 2024, the court ordered the remaining claim for copyright infringement in *Kadrey* to be consolidated with the claims in *Huckabee v. Bloomberg*.[395] Whether this legal battle will be successful for the plaintiff authors remains to be seen, but the discovery process thus far has swayed the public in their favor. To the horror of some Meta employees, the discovery process has uncovered the company's use of pirated material from Library Genesis (LibGen) to train its AI language models. [396] Meta employees' use and possible dissemination of large amounts of copyrighted material through controversial "shadow libraries" like LibGen could have lasting effects on its public image.[397] In addition to this public relations blow, a potential ruling in favor of the plaintiffs would force Meta and other technology companies to reevaluate their AI training methods to avoid further litigation supported by this precedent. [398] Some intellectual property lawyers believe the plaintiffs' remaining direct copyright infringement claim may be successful now that their complaint has been amended to reflect LLaMA's use of known datasets which include the plaintiffs' books.[399]

Tags: artificial intelligence, AI, copyright, AI training, Sarah Silverman, copyright infringement, Digital Millennium Copyright Act, DMCA, Meta, Large Language Model Meta AI, LLaMA, unjust enrichment, negligence, fair use

---

[392] *Id*.

[393] Knowing Machines, *Kadrey v. Meta Legal Explainer* (Nov. 27, 2023), https://knowingmachines.org/knowing-legal-machines/legal-explainer/cases/kadrey-v-meta

[394] *Is the Copyright Threat to Generative AI Overhyped? Implications of Kadrey v. Meta,* FARELLA, BRAUN & MARTEL: INSIGHT PUBLICATIONS (Jan. 13, 2024), https://www.fbm.com/publications/is-the-copyright-threat-to-generative-ai-overhyped-implications-of-kadrey-v-meta/.

[395] *Kadrey v. Meta Case Tracker, supra* note 341.

[396] Kate Knibbs, *Meta Secretly Trained Its AI on a Notorious Piracy Database, Newly Unredacted Court Docs Reveal,* WIRED (Jan. 9, 2025), https://www.wired.com/story/new-documents-unredacted-meta-copyright-ai-lawsuit/.

[397] *Id.*

[398] *Id.*

[399] FARELLA, BRAUN & MARTEL: INSIGHT PUBLICATIONS, *supra* note 385.

J.   Nazemian and Dubus v. NVIDIA Corporation. Abdi Nazemian et al v. NVIDIA Corporation, No. 3:24-cv-01454 and Dubus, et al. v. NVIDIA, Inc., No. 3:24-cv-02655-VC (N.D. Cal.

This is a class-action suit filed by two groups of authors: Abdi Nazemian, Brian Keene, and Stewart O'Nan (the Navemian case) and Andrew Dubus III & Susan Orlean (Dubus) against NVIDIA Corporation, a US technology company that is a dominant supplier of AI hardware and software. The Plaintiffs allege that NVIDIA used and copied their books to train its large language model without permission. The big question of this case is whether usage of these copyrighted works in training can be considered under the defense of fair use, which would then absolve NVIDIA from any copyright infringement liability.

Tags: fair use, copyright infringement, NVIDIA

K. The Newspaper Cases against Microsoft

This is a consolidated case with *New York Times v. Microsoft*,[400] *Daily News v. Microsoft* [401]and *Center for Investigative Reporting v. Microsoft*.[402] Microsoft, the defendant in each of these actions, is a technology company and financial backer of OpenAI.[403] OpenAI is an AI research and development company.[404] The *New York Times* ("the Times") is a newspaper organization which covers domestic and national current events based in New York City.[405] The Times filed a complaint on December 27, 2023, that OpenAI fed the software information from the Times to train it without compensation nor the Times' permission to do so.[406] The Times claimed that OpenAI's usage of their materials may operate as a substitute for a Times subscription.[407]

---

[400] The New York Times v. Microsoft, et. al., No. 1:23-cv-11195 (S.D.N.Y.. filed Dec. 27, 2023).

[401] Daily News, LP et al. v. Microsoft Corp. et. al., No. 1:24-cv-03285 (S.D.N.Y. filed April 30, 2024).

[402] The Ctr. for Investigative Reporting, Inc. v. Microsoft Corp. et. al., No. 1:24-cv-04872 (S.D.N.Y. filed June 27, 2024)

[403] *Microsoft and OpenAI evolve partnership to drive the next phase of AI*, MICROSOFT, https://blogs.microsoft.com/blog/2025/01/21/microsoft-and-openai-evolve-partnership-to-drive-the-next-phase-of-ai/#:~:text=Microsoft%20remains%20a%20major%20investor,from%20their%20growth%20in%20valuation (last visited Feb. 17, 2025).

[404] *About*, OPENAI, https://openai.com/about/ (last visited Feb. 17, 2025).

[405] *Company*, THE N.Y. TIMES CO.., https://www.nytco.com/company/ (last visited Feb. 17, 2025).

[406] The New York Times v. Microsoft, No. 1:23-cv-11195 at *15

[407] The New York Times v. Microsoft, No. 1:23-cv-11195-SHS (filed Feb. 26, 2024)

Daily News, L.P, doing business as the *New York Daily News*, is a newspaper publishing company.[408] The Daily News' complaint includes seven other news organizations as plaintiffs, collectively known as "the Publishers." [409] The Publishers asserts that Defendants used their copyrighted materials such as articles in order to train their software without permission or compensation.[410] The Publishers bring an additional New York state-law trademark dilution claim.[411]

The Center for Investigative Reporting ("CIR") is a non-profit which seeks to publish pieces of investigative journalism, primarily through their radio show and podcast "Reveal".[412] CIR claims OpenAI copied, abridged, and displayed their content without permission, authorization, or compensation from Microsoft.[413] Due to the similarities of their claims, these three cases consolidated, becoming "the Newspaper Cases" on June 13, 2024. [414] The newspaper organizations argue that this use of copyrighted material without permission or payment in combination with many people's use of OpenAI as a substitute for the copyrighted materials within the articles.[415]

OpenAI argues that the data used to train their software is covered under fair use, and that large language models work with the data in a transformative way.[416] OpenAI asserts this information is publicly accessible and that it violates copyright law to "monopolize facts."[417] Key questions include whether OpenAI used materials from these organizations to train their software in a way that infringed on copyrighted materials.

Other news organizations such as The Associated Press, The Atlantic, Vox Media, FT Group, Axel Springer, Dotdash Meredith, NewsCorp, Guardian, Schibsted, Guardian, Axios, Future, Hearst, FT, Reuters, Fortune, USA Today Network, Conde Nast, Time, Informa, Financial Times, Le Monde, Prisa Media,

---

[408] *About Us*, DAILY NEWS, https://www.nydailynews.com/about-us/ (last visited Feb. 17, 2025).

[409] *Daily News, LP et al. v. Microsoft Corp. et. al.*, No. 1:24-cv-03285, at 1

[410] *Id*. at 2

[411] *Id*. at 93

[412] *About Us*, REVEAL, https://revealnews.org/about-us/ (last visited Feb. 17, 2025).

[413] *Ctr. for Investigative Journalism v. Microsoft,* No. 1:24-cv-04872 at *2-3

[414] Notice of Motion and Motion to Consolidate, The New York Times v. Microsoft, et. al and Daily News, LP et al. v. Microsoft Corp. et. al, No. 1:23-cv-11195-SHS (S.D.N.Y. 2024)

[415] *Id*.

[416] *Id*.

[417] Notice of Motion Oral Argument Requested, N.Y. Times Co. v, Microsoft Corp. et al., No. 1:23-cv-11195 (S.D.N.Y. 2023).

and Shutterstock have entered into licensing agreements with OpenAI or Microsoft to permit the usage of their articles to train the AI.[418]

Seeing as OpenAI is entering into licensing deals with 34 newspaper publishing organizations and counting, it is likely that this is an element of their litigation strategy. By entering into these arrangements, it weakens the Times and others' positions that OpenAI stole their content without consent.

Tags: newspaper, OpenAi, copyright, license, Microsoft, CIR, news, fair use, information.

L.   OpenAI ChatGPT Litigation, [In re OPENAI CHATGPT Litigation, No. 3:23-cv-03223-AMO (N.D. Cal. July 12, 2024).] [419]

This is a consolidation case, filed in the Northern District of California that involves three plaintiff groups that consist of fiction and nonfiction authors Paul Tremblay, Sarah Silverman, Richard Kadrey, and Christopher Golden,[OBJ][420] as well as Michael Chabon, Laura Lippman, Andrew Sean Greer, Ta-Nehisi Coates, David Henry Hwang, Rachel Louise Snyder, Laura Lippman, Ayelet Waldman, Junot Díaz, Jacqueline Woodson.[421] They filed on behalf of themselves and a putative class [OBJ][422] (*Tremblay v. OpenAI*,[OBJ][423][OBJ] and *Chabon v. OpenAI).*[OBJ][424] Each plaintiff group has filed a complaint against defendant OpenAI, alleging copyright infringement, DMCA violations, vicarious copyright infringement, and torts associated with OpenAI's GPT models and ChatGPT service.[425]

---

[418] Cecily Mauran, *All the Media Companies that have Licensing Deals with OpenAI (so far)*, MASHABLE: TECH (June 21, 2024), https://mashable.com/article/all-the-media-companies-that-have-licensing-deals-with-openai-so-far; Charlotte Tobitt, *Who's suing AI and who's signing: 14 publishers join lawsuit against start-up Cohere*, PRESS GAZETTE (Feb. 14, 2025), https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/.

[419] Natalia E. Walcott, Tulane University Law School, Class of 2025.

[420] *OpenAI ChatGPT Litigation,* BAKERHOSTETLER, https://www.bakerlaw.com/openai-chatgpt-litigation/ (last visited Feb. 8, 2025). This set of Plaintiffs are represented by Joseph Saveri Law Firm, LLP and Matthew Butterick.

[421] *Id.* This group of Plaintiffs are represented by Ventura Hersey & Muller, LLP and Caferty Clobes Meriwether & Sprengel LLP.

[422] *Id.*

[423] Tremblay v. OpenAI, Inc., No. 3:23-cv-03223 (N.D. Cal. filed June 28, 2023); Silverman v. OpenAI, Inc., No. 3:23-cv-03416 (N.D. Cal. filed July 7, 2023).

[424] Chabon v. OpenAI, Inc., No. 3:23-cv-04625 (N.D. Cal. filed Sept. 8, 2023).

[425] BAKERHOSTETLER, *supra* note 420. (Noting that the defendant in the three cases is OpenAI and includes OpenAI, Inc., OpenAI, L.P., OpenAI GP, L.L.C., OpenAI OpCo, L.L.C., OpenAI Startup Fund GP I, L.L.C., OpenAI Startup Fund I, L.P., and OpenAI Startup Fund Management, LLC. Defendants are represented by Morrison & Foerster LLP and Latham & Watkins LLP).

Tremblay's initial complaint discussed OpenAI's model.[426] OpenAI creates its large language models ("LLM") by copying a large amount of text from different types of sources.[427] The LLMs then take these copies, absorb the text, and create work that seems to be that written by a human.[428] In this case, the plaintiffs, Paul Tremblay and Mona Awad, alleged that OpenAI used plaintiffs' books without their knowledge or permission, in the training dataset for ChatGPT. [429] They discovered this when they prompted ChatGPT to generate accurate summaries of their copyrighted books.[430] On a larger scale, they discuss OpenAI's copying of many books in their data sets, respectively Books1 and Books2, which were used to train their GPT-3, and subsequent, models.[431] It is believed that Books1 and Books2 have over 350,000 titles combined, though OpenAI has not disclosed the names of those titles. [432] They alleged that the defendants are "benefit[ting] commercial[ly] and profit[ing] richly" by using their copyrighted materials through the ChatGPT platform. [433] Since the defendants have not disclosed the names of the titles used in their datasets, the size and members of the class are unknown.[434] Due to this, the plaintiffs are open to representing all potential parties in this case.[435]

The allegations in *Tremblay* are the following:
- Copyright infringement
  - Direct copyright infringement
  - Vicarious copyright infringement
- DMCA
  - Violation of the Digital Millennium Copyright Act through their removal of copyright-management information
- Other Counts
  - Unfair competition
  - Negligence
  - Unjust enrichment.[436]

---

[426] Complaint at 4, *Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03223 (N.D. Cal. filed June 28, 2023), ECF No. 1.

[427] *Id.*

[428] *Id. at 4-5.*

[429] *Id.* at 2, 8. (stating that Plaintiffs Paul Tremblay and Mona Awad are writers. Tremblay owns several copyrighted books, including *The Cabin at the End of the World*. Awad also owns several copyrighted books, including *13 Ways of Looking at a Fat Girl and Bunny*).

[430] *Id. a*t 1.

[431] *Id. a*t 6.

[432] *Id.* (noting that Book1 is believed to have about 63,000 titles, and Books2 is to have around 294,000 titles).

[433] *Id. a*t 1.

[434] *Id. a*t 8-10.

[435] *Id.*

[436] *Id. a*t 10-15.

The plaintiffs welcome the lawsuit becoming a class action and are seeking statutory and other damages, permanent injunctive relief, attorney's fees and other costs associated with the lawsuit.[437]

Silverman's initial complaint discusses OpenAI's use of the Plaintiffs' copyrighted materials in their training data sets to condition their AI models to create their LLMs.[438] The plaintiffs in this lawsuit are authors Sarah Silverman, Christopher Golden, and Richard Kadrey.[439] This lawsuit discussed the data sets OpenAI used to train their models, stating that the Defendants did not disclose the names of the copyrighted material in the sets.[440] The Plaintiffs believe that their work is included in the data sets, and that, when prompted, ChatGPT created a summary of their books, using similar text.[441] They are concerned about the inaccuracies that can occur with these models, as it could cause a false representation of the contents of their individual works.[442]

The allegations in *Silverman* are the following:

- Copyright infringement
  - Direct copyright infringement
  - Vicarious copyright infringement

- DMCA
  - Violation of the Digital Millennium Copyright Act through their removal of copyright-management information

- Other Counts
  - Unfair competition
  - Negligence
  - Unjust enrichment.[443]
  -

Similar to *Tremblay,* the plaintiffs are open to making this a class action suit and are seeking attorney's fees, statutory and other damages, permanent injunctive relief, and other lawsuit-related costs.[444]

---

[437] *Id. a*t 15.

[438] Complaint at 4-5, Silverman v. OpenAI, Inc., No. 3:23-cv-03416 (N.D. Cal. filed July 7, 2023), ECF No. 1.

[439] *Id.* at 2-3. Plaintiffs Sarah Silverman, Christopher Golden and Richard Kadrey are all writers. Silverman owns several copyrighted books, including *The Bedwetter*. Golden owns several copyrighted books, including *Ararat*. Kadrey also owns several copyrighted books, including *Sandman Slim*.

[440] *Id.* at 6.

[441] *Id.* at 8.

[442] *Id.*

[443] *Id.* at 10-15.

[444] *Id.* at 15-16.

The final case in this consolidated lawsuit is *Chabon v. OpenAI.*[445] The initial complaint was filed by award-winning authors Michael Chabon, David Henry Hwang, Matther Klam, Rachel Louise Snyder, and Ayelet Waldman.[446] They made allegations similar to that of *Tremblay* and *Silverman,* alleging that OpenAI did not get the authors' permission to use their books in OpenAI's training dataset.[447] Plaintiffs believe ChatGPT and the model's use of their copyrighted material is an unauthorized use of derivative work.[448] They argued that OpenAI's acknowledgement of using various sources for their training models implied that their work was one of those pieces used.[449] Plaintiffs believe that OpenAI's rationale for doing this stems from their need to use sources of high quality to make sure that their model's outputs reflect work of the same or similar caliber.[450] Similar to *Tremblay and Silverman*, they discussed class allegations, due to the fact that other members of the class are undisclosed in the Defendants' data sets.[451]

The following counts alleged in *Chabon* are the following:

- Copyright infringement
    - Direct copyright infringement
    - Vicarious copyright infringement

- DMCA

---

[445] Chabon v. OpenAI, Inc., No. 3:23-cv-04625 (N.D. Cal. filed Sept. 8, 2023).

[446] Complaint at 2-3, Chabon v. OpenAI, Inc., No. 3:23-cv-04625 (N.D. Cal. filed Sept. 8, 2023), ECF No. 1. (noting that Plaintiffs Michael Chabon, David Henry Hwang, Matthew Klam, Rachel Louise Snyder, and Ayelet Waldman are writers. Chabon owns several copyrighted books, "including but not limited to *The Mysteries of Pittsburgh, Wonder Boys, The Amazing Adventures of Kavalier & Clay, the Yiddish Policemen's Union, Gentlemen of the Road, Telegraph Avenue, Fight of the Century, Kingdom of Olive and Ash,* and *Moonglow.*" Hwang owns several copyrighted books, "including but not limited to *M. Butterfly, Chinglish, Yellow Face, the Dance and the Railroad,* and *FOB.*" Klam owns several copyrighted books, "including but not limited to, *Who is Rich?,* and *Sam the Cat and Other Stories.*" Snyder owns several copyrighted books, "including but not limited to, *Women We Buried, Women We Burned, No Visible Bruises – What We Don't Know About Domestic Violence Can Kill Us, What We've Lost is Nothing,* and *Fugitive Denim: A Moving Story of People and Pants in the Borderless World of Global Trade.*" Waldman owns several copyrighted books, "including but not limited to, *Love and other Impossible Pursuits, Red Hook Road, Love and Treasure, Bad Mother, Daughter's Keeper, A Really Good Day, Fight of the Century,* and *Kingdom of Olives* and *Ash.*").

[447] *Id.* at 8-11. *See also* BAKERHOSTETLER, *supra* note 433.

[448] *Chabon v. OpenAI,* BAKERHOSTETLER, https://www.bakerlaw.com/chabon-v-openai/ (last visited Feb. 8, 2025).

[449] Complaint, Chabon, *supra* note 445 at 6.

[450] *Id.* at 5-7.

[451] *Id.* at 8-10.

         ○   Violation of the Digital Millennium Copyright Act through their removal of copyright-management information

- Other Counts
  - ○ Unfair competition
  - ○ Negligence
  - ○ Unjust enrichment.[452]

On November 7, 2023, the Northern District of California consolidated *Tremblay*, *Silverman*, and *Chabon*.[453] The consolidated group went on to file an amended complaint on March 13, 2023, alleging the same counts that were made in the three separate cases.[454] Open AI filed a motion to dismiss Count II in the Amended Complaint, alleging that the Plaintiffs failed to state a claim for their Unfair Competition Law (UCL) claim.[455] The court granted the Motion to Dismiss Count II, denying them the ability to leave to amend.[456]

There have not been many substantial updates on the case.[457] The case is currently in the discovery phase, with both parties filing a joint brief discussing their discovery concerns.[458] Plaintiffs alleged that OpenAI refuses to produce documentation about their language models, while OpenAI says that the information in those documents is irrelevant because ChatGPT does not use in-development models.[459] We will have to see what the Court decides in this matter.

Tags: OpenAI, Large Language Models ("LLMS"), Chabon, Silverman, Tremblay, Digital Millennium Copyright Act, OpenAI GPT models, ChatGPT, GPT models, direct copyright infringement, copyright infringement, unfair competition, negligence, and unjust enrichment

M. Thomson Reuters v. ROSS, 694 F. Supp. 3d 467 (D. Del. 2023).[460]

---

[452] *Id.* at 10-14.

[453] Pretrial Order at 1, In re OPENAI CHATGPT Litigation, No. 3:23-cv-03223-AMO (N.D. Cal. filed July 12, 2024), ECF No. 74.

[454] Amended Complaint at 1, In re OPENAI CHATGPT Litigation, No. 3:23-cv-03223-AMO (N.D. Cal. filed July 12, 2024), ECF No. 120.

[455] Defendant's Notice of Motion, Motion to Dismiss FirstConsolidated Amended Complaint, and Memorandum of Points and Authorities in Support Thereof at 4-7, In re OPENAI CHATGPT Litigation, No. 3:23-cv-03223-AMO (N.D. Cal. filed July 12, 2024), ECF No. 122.

[456] Order Granting Motion to Dismiss the UCL Claim at 5, In re OPENAI CHATGPT Litigation, No. 3:23-cv-03223-AMO (N.D. Cal. filed July 12, 2024), ECF No. 162.

[457] *AI Infringement Case Updates: December 16, 2024,* MCKOOL SMITH, (Dec. 16, 2024), https://www.mckoolsmith.com/newsroom-ailitigation-3.

[458] *Id.*

[459] *Id.*

[460] Jon Evan Smelley, Tulane Law School, Expected Graduation 2025.

The issue in this case is whether ROSS's use of Westlaw's content constituted fair use or whether it was unauthorized and constituted as direct infringement.[461] This was one of the first AI cases filed.[462] The judge over the case, Judge Stephanos Bibas, stated that it would not be his place to rule as a matter of law on if ROSS's use of Westlaw's content was fair use.[463] This issue will be left for the jury to decide on a "headnote-by-headnote basis."[464] Although the trial was previously set for August 23, 2024, it has been pushed to May 12, 2025; however, due to the costs of litigation, ROSS shut down their platform in January of 2021.[465] The shutdown does not automatically absolve ROSS of potential copyright infringement liability.[466] If the court finds that ROSS unlawfully used Westlaw content prior to shutting down, damages may still be awarded based on the impact of that alleged infringement.[467] On the other hand, ROSS could argue that the platform's shutdown mitigates any potential harm to Thomson Reuters, potentially reducing or eliminating damages.[468] Additionally, the shutdown underscores a broader question in the case—whether AI companies can use copyrighted legal databases for training purposes without facing legal consequences.[469]

This case has been going for nearly five years, beginning in May 2020 when Thomson Reuters sued ROSS Intelligence for allegedly copying Westlaw content to train its AI-based legal research platform.[470] Thomson Reuters contends that ROSS's use of its proprietary legal materials constitutes copyright infringement,

---

[461] Aaron P. Rubin, *Reuters v. ROSS: District Court First to Consider Whether Training Generative AI Model Is Fair Use*, MORRISON FOERSTER (Oct. 20, 2023), https://www.mofo.com/resources/insights/231020-reuters-v-ross-district-court; Aaron Moss, *First AI Copyright Trial Starts This Week: What to Know*, COPYRIGHT LATELY (Aug. 19, 2024), https://copyrightlately.com/first-ai-copyright-trial-starts-westlaw/.

[462] *Id.*

[463] Rubin, *supra* note 461.

[464] *Id.*

[465] *Judge Bibas sets trial for May 12, 2025, but hopes for earlier date in Thomson Reuters v. ROSS Intelligence,* CHAT GPT IS EATING THE WORLD (Nov. 4, 2024), https://chatgptiseatingtheworld.com/2024/11/04/judge-bibas-sets-trial-for-may-12-2025-but-hopes-for-earlier-date-in-thomson-reuters-v-ross-intelligence/; The Founders, *Announcement,* ROSS INTELLIGENCE (Dec. 11, 2020), https://blog.rossintelligence.com/post/announcement.

[466] The Founders, *supra* note 365; Reuters, *Jury to decide on whether AI was trained on copied material in one of first major AI training trials,* EURO NEWS (Sep. 26, 2023, 2:16 PM GMT), https://www.euronews.com/next/2023/09/26/jury-to-decide-on-whether-ai-was-trained-on-copied-material-in-one-of-first-major-ai-train; *Thomson Reuters v. ROSS,* BAKER HOSTETLER, https://www.bakerlaw.com/thomson-reuters-v-ross/ (last visited Feb. 12, 2025).

[467] *Id.*

[468] *Id.*

[469] *Id.*

[470] BAKERHOSTETLER, *supra* note 466.

while ROSS argues that its use falls under fair use as it was transformative rather than merely duplicative.[471] Over the years, the case has seen multiple legal motions, hearings, and counterclaims, culminating in a renewed debate over copyright infringement and fair use.[472]

The case commenced with ROSS filing a motion to dismiss in July 2020, which was ultimately denied in March 2021.[473] The litigation escalated when both parties filed motions for summary judgment in December 2022, with Thomson Reuters seeking to establish copyright infringement and ROSS defending itself under the fair use doctrine.[474] Throughout 2023, the court denied motions for summary judgment and scheduled the trial for August 2024.[475] In 2024, the case evolved with discussions on whether Westlaw's headnotes were copyrightable, leading to a filtration hearing request by ROSS, which the court later declined.[476]

By the latter half of 2024, both parties renewed their summary judgment motions on copyright infringement and fair use, submitting arguments throughout October.[477] On December 5, 2024, the court held a hearing on these motions, with a ruling still pending.[478] The outcome of this case could significantly impact the future of AI-driven legal research and the scope of copyright protections in the digital age.[479] This case looks like it will be the first case that will decide whether fair use will prevail when it comes to AI.[480]

Tags: AI, Copyright infringement, Tortious Interference, No Fair Use, Fair Use, Antitrust, Anticompetition, Summary Judgement, Filtration Hearing

N.   Voice Cases

There are a number of AI-related voice cases where unauthorized use of someone's voice is the issue. Voice actors have brought a class action suit against LOVO for using their voices to create generated voices without their consent.[481] ElevenLabs is also getting sued for the same kind of conduct.

---

[471] *Id.*
[472] *Id.*
[473] *Id.*
[474] *Id.*
[475] *Id.*
[476] *Id.*
[477] *Id.*
[478] *Id.*
[479] Rubin, *supra* note 461; Moss, *supra* note 461.
[480] *Id.*
[481] Jane Wester, *Control of Their Careers Has Been Hijacked': Voice Actors File Proposed Class Action Against AI Company*, ALM LAW, https://www.law.com/2024/05/20/control-of-their-careers-has-been-hijacked-voice-actors-file-proposed-class-action-against-ai-company/?slreturn=2025030261957 (last visited March 1, 2025).

IV. SECONDARY SOURCES RELATED TO THE CASES[482]

The cases are one aspect of what we should keep our eye on. We also wanted to understand the themes we are seeing, and how people are writing about the cases at the progress through the system. How can we understand all of these cases in a coherent way? We see a number of resources trying to help us with that as well. Using secondary sources, we wanted to know how the news was reporting the causes of action. What are the themes? The two key themes are direct infringement and DMCA claims.

AI companies accessed copyrighted works and made copies for training the AI models.[483] This unauthorized copying comes in additional steps in the process including "weight folders" and producing substantially similar outputs that are also infringing.

Copyright owners have filed lawsuits in U.S federal court against AI companies alleging direct copyright infringement for improper use of copyrighted material.[484] Many of the cases assert unauthorized removal or alteration of copyright management information as well as claiming infringing derivative works.[485] While many of these claims have been dismissed, none of the unauthorized use of works claims have been dismissed which are the focus of these cases.[486] While AI cases are still in the early stages, practitioners have already shown a tendency for courts to side with plaintiffs.[487]

New York Times Attorney, Ian Crosby, described AI's impact as "substantial."[488] Users are now able to search questions and receive responses.[489] This point is pivotal in establishing a clear-cut copyright infringement case.[490] Publishers and authors assert that the data that drives ChatGPT is mainly supported by copyrighted content used without consent or authorization.[491] The publisher's legal team argues that ChatGPT profits from taking, processing, and

---

[482] Tynesea Watts, Tulane Law School, Expected Graduation 2025.

[483] David M. McIntosh et al., *An End-of-Year Update to LLP*, ROPES GRAY (Dec 17, 2024), https://www.ropesgray.com/en/insights/alerts/2024/12/an-end-of-year-update-to-the-current-state-of-ai-related-copyright-litigation.

[484] Keith Kupferschimd, *Insights from Court Orders in AI Copyright Infringement Cases*, COPYRIGHT ALLIANCE (Dec 12, 2024), https://copyrightalliance.org/ai-copyright-infringement-cases-insights/.

[485] *Id*. at 282.

[486] *Id*.

[487] *Id*.

[488] Complaint, The New York Times Co. v. Microsoft Corp. et al., 1:23-cv-11195 (S.D.N.Y. 2023).

[489] Bobby Allyn, *'The New York Times' takes Open AI to court. ChatGPT's future could be on the line*, NPR (Feb. 9, 2025), https://www.npr.org/2025/01/14/nx-s1-5258952/new-york-times-openai-microsoft

[490] *Id*.

[491] *Id*.

recreating journalistic material without payment or authorization.[492] In response, AI companies are clinging to the fair use doctrine as defense by claiming that AI tools.[493]

A victory for publishers and authors would grant them more control over their work and its use. However, a loss for AI companies would require that the AI companies destroy their current dataset and recreate a new dataset with work they are authorized to use.[494] In contrast, the weight of the fair use defense in copyright cases hangs in the balance. A win for AI companies would constitute an expansion of the fair use defense, while giving creators less control of who can access their work and how their work can be used.

The Thomson v. ROSS case sits at the precipice of this controversy[495]. In Thomson v. ROSS, a federal district court found that the headnotes and a key numbering system that ROSS AI ("ROSS') copied from Thomson Reuters Westlaw ("Thomson") satisfied the legal requirement for originality. Subsequently, the court decided that ROSS copied copyrightable material from Thomson.

This decision received strong criticism since its ruling.[496] Critiques questioned the originality of the material in question by claiming that the Judge mischaracterized the merger doctrine.[497] The court labeled near- verbatim summaries as original work by ruling that the headnotes are original individually and as a compilation and by ruling that the key number system was original as and protected by copyright as a compilation.[498]

In US copyright law "originality" specifically refers to work that has a modicum of creativity that our society would want to protect and encourage.[499] Additionally, in US copyright law, facts are not copyrightable, but sometimes the

---

[492] *Id.*

[493] *Delaware District Court Ruling Raises Critical Fair Use Challenges For AI Companies: Implications For Copyright Compliance and Competitive Practices,* DYKEMA *(Feb. 15, 2025), https://www.dykema.com/news-insights/delaware-district-court-ruling-raises-critical-fair-use-challenges-for-ai-companies-implications-for-copyright-compliance-and-competitive-practices.html*

[494] *Id.*

[495] Joseph A. Meckes, *Court: Training AI Model Based on Copyrighted Data Is Not Fair Use as a Matter of Law*, THE NAT. L. REV.: GLOBAL IP AND TECH L. BLOG (Feb. 11, 2025), https://natlawreview.com/article/court-training-ai-model-based-copyrighted-data-not-fair-use-matter-law.

[496] *Thomas Reuters v. ROSS: The First AI Fair Use Ruling Fails to Persuade,* AUTHORS ALLIANCE, *(Feb. 13, 2025),* https://www.authorsalliance.org/2025/02/13/thomson-reuters-v-ross-the-first-ai-fair-use-ruling-fails-to-persuade/

[497] *Id.*

[498] *Id.*

[499] Paul Matenear, *Copy That! Copyright Infringement in Branding,* AXLEY ATTORNEYS (Feb 13, 2025), https://www.axley.com/publication_article/copy-that-infringement-in-branding/.

expression of facts can be.[500] The doctrine of merger also protects our freedom to reference facts or an idea.[501] US courts have recognized that there are a limited number of ways that one can restate facts, and therefore those expressions don't qualify as "original." [502]   The Doctrine of Merger recognizes that certain expressions "merge" with facts and thereafter become unprotectable.[503]

Critics assert that Thomson's key numbering system and headnotes lack the modicum of creativity required for originality because ROSS copied near-verbatim summaries of facts. [504] Moreover, critics assert that the judge mischaracterized the doctrine of merger and ruled that it doesn't apply in this case because he believes that there are many ways to express points of judicial law from judicial opinion. [505] Critics claim that this reasoning shows a flawed understanding of the doctrine of merger because the very nature of language allows for an expression to be said in multiple ways, but the doctrine of merger applies when the number of ways to retell an expression is limited – especially when referring to legal expressions and terms of art.[506]

The most notable critique of this case is the fair use ruling. The court ruled that ROSS's use of Thomson's headnotes and key numbering system was not fair use because it did not satisfy two of the four factors in the fair use analysis.[507] The District Court decided that ROSS's fair use argument did not prove transformative use under the first factor because of ROSS's strictly commercial use. [508] Subsequently, the court found that ROSS's use failed the fourth factor because ROSS was unable to show lack of a potential market for Thomson to sell its content as AI training data and that such market would be unaffected by ROSS's actions.[509]

The fair use ruling attracted a significant amount of criticism. In order to decide if the use of the headnotes and the key numbering system counted as transformative use under the first factor of the analysis, the court had to assess whether intermediate copying case law supported ROSS's use.[510] The court asserts that the copying of the head notes and the key number system counts as copyright infringement because an intermediate copy with copyright restricted

---

[500]Molly Shaffer Van Houweling, *The Freedom to Extract in Copyright Law*, 103 N.C. L. Rev. 447-448 (2025).

[501] *Id.*

[502] *Id.*

[503] *Id.*

[504] Authors Alliance, *supra* note 496.

[505] *Id.*

[506] *Id.*

[507] Mark Humphry, *Court Unplugs AI Fair Use Defense, But Context is Key*, Mitchell Silberberg & Knupp LLP   (Feb 14, 2025), https://www.msk.com/newsroom-alerts-ai-fair-use-decision

[508] *Id.*

[509] *Id.*

[510] Author's Alliance, *Supra* note 499.

material authored by Thomson exists, even though the exact content from Thomson was not shown to the public.[511] According to this court, intermediate copying only supports fair use for computer codes.[512] Critics highlight the fact that the judge contradicts their reasoning in this opinion because under the third factor of the fair use analysis the judge states that only content available to the public when analyzing copyright infringement.[513]

Critics also note the court's discussion regarding intermediate copying by asserting that case law supports fair use when the material in question is not shared to the public and there is no intention to share the content with the public.[514] Lasty, many critics have challenged the court's analysis of the fourth factor by stating that the mere creation of the derivative work implies that a market for this work already exists.[515] The judge insinuates that by using Thomson's content, ROSS created an AI training data market that Thomson should have control of as the copyright holder.[516] Critics argue the reasoning used in this analysis is not supported by case law.[517] Rather, they say to properly analyze the fourth factor copyright holders must prove the existence, or likelihood of developing, licensing market, before arguing a secondary use serves as "market substitute."[518]

V. LICENSING IN THE SHADOW OF LAWSUITS[519]

A. Licensing Deals

We are seeing a trend – licensing deals with major AI players, while the lawsuits continue. Here are some key news stories and resources that are keeping track. Here are three lists to get you started.

- Sara Guaglione, "2024 in review: A Timeline of the major deals between publishers and AI companies, Digiday, December 27, 2024. https://digiday.com/media/2024-in-review-a-timeline-of-the-major-deals-between-publishers-and-ai-companies/

   She reports that 2024 saw more formalized relationships with a wave of licensing agreements to use publishers' content to train LLMs.

---

[511] *Id.*

[512] *Id.*

[513] *Id.*

[514] *Fair Warning: AI's First Copyright Fair Use Ruling, Thomson Reuters v. Ross*, HUNTON (Feb 19, 2025), https://www.hunton.com/insights/legal/fair-warning-ais-first-copyright-fair-use-ruling-thomson-reuters-v-ross

[515] *Id.*

[516] *Id.*

[517] *Id.*

[518] *Authors Alliance Submits Amicus Brief In Sedlik V. Drachenberg*, AUTHOR'S ALLIANCE (Feb 15, 2025), https://www.authorsalliance.org/2024/12/23/authors-alliance-submits-amicus-brief-in-sedlik-v-drachenberg/

[519] Elizabeth Townsend Gard

Publishers also gain attribution that their materials are being used, including in chatbots and search platforms. She takes us through a timeline of when and the details of agreements that have been made.

● Thani Shamsi, "Ultimate List of Data Licensing Deal for AI, October 28, 2024, Monda, https://www.monda.ai/blog/ultimate-list-of-data-licensing-deals-for-ai.

This article approaches the question from what organizations have made deal with AI companies, e.t. Reddit, Shutterstock, etc. It profies ten companies.

● Charlotte Tobitt, "Who's Suing AI and who's signing: Major deals with Google, OpenAI and Mistral start year," PressGazette, January 25, 2025, https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/.

This article lists the nine cases where media is suing AI companies, and with it summaries of each case. It also lists the news publishers that have agreements with AI companies, listed by the publishers. The author promises that the page will be updated as legal cases proceed and new deals are made. So, this is one to bookmark.

So, who has made agreements so far?

Agreements with OpenAI
Le Monde
Prisa
Financial Times
DotDash Meredith
News Corp.
The Atlantic
Vox Media
Time
Conde Nast
Hearst
Future
Associated Press
Shutterstock
Axios, including funding four newsrooms with Axios.[520]

Agreements with Microsoft
Axel Springer

---

[520] Daniel Konstantinovic, *OpenAI to Fund Four Axios Newsrooms in Content Licensing Deal*, EMARKETER (Jan. 15, 2025), https://www.emarketer.com/content/openai-fund-four-axios-newsrooms-content-licensing-deal

Informa
FT
Reuters
Hearst
USA Today Network
Fortune

Agreements with Perplexity[521]
Time
Der Spiegel
Fortune
Entrepreneur
The Texas Tribune
ADWEEK
Blavity
DPReview
Gear Patrol
The Independent
Lee Enterprises
Los Angeles Times
MediaLab
Mexico News Daily
Minkabu Infonoid
NewsPicks
Prisa Media
RLT Germany grands stern and nvt
World History Encyclopedia

Agreements with ProRata
Financial Times
Axel Springer
The Atlantic
Fortune
DMG media
Sky News
Guardian
Prospect

B.   *A New Lawsuit: Advance Local Media v. Cohere, 1:25-cv-01305 (SDNY)*

---

[521] *Welcoming more global news organizations to Perplexity's Publishers' Program*, ANDERSEN (Dec. 5, 2024), https://www.perplexity.ai/hub/blog/perplexity-expands-publisher-program-with-15-new-media-partners#.

On February 13, 2025, a collection of news publishers that allege copyright infringement and false attribution by Cohere in the building of their AI. What is interesting about this case is that the complaint details the including that there is availability of licensing as an alternative to copyright infringement. These news groups that are part of the lawsuit include Conde Nast (who has licensed with OpenAI), The Atlantic (who has licensed with ProRata and OpenAI), Axel Spring (who has licensed with ProRata), Forbes, Guardian, Los Angeles Times, Politico, The Republican, Plain Dealer, Toronto Star Newspapers, McClatchy (with 30 daily newspapers), Newsday, and Vox (who has licensed with OpenAI). The complaint begins, "This is a lawsuit to protect journalism from systematic copyright and trademark infringement. Rather than create its own content, Cohere takes the creative output of Publishers, some of the largest, most enduring, and most important news, magazine, and digital publishers in the United States and around the world. Without permission or compensation, Cohere uses scraped copies of our articles, through training, real-time use, and in outputs, to power its artificial intelligence ("AI") service, which in turn competes with Publisher offerings and the emerging market for AI licensing. Not content with just stealing our works, Cohere also blatantly manufactures fake pieces and attributes them to us, misleading the public and tarnishing our brands."[522] So, the licensing deals with AI companies have led to a market that, according to the complaint does not justify fair use of using news materials.

And who is Cohere? According to the complaint, it's a $5 billion Canadian company who is training on unlicensed materials and the outputs include "full verbatim copies, substantial excerpts, and substantive summaries."[523] It prides itself on its news reporting function, using the plaintiffs' materials unlicensed.

Included in the complaint are the licensing deals the plaintiffs have both with traditional platforms like Yahoo!. Apple News, NewsBreak, and Ancestry.com, Lexis Nexis, but also AI companies. They own not only copyrights that are being infringed, but also trademarks. What is interesting is that the complaint in many ways will mimic the summary judgement in ROSS, showing that the summaries are similar to the original articles. So, they are alleging infringement not only in the training but also the outputs. The 64-page complaint has many examples with graphics to show the copying in various forms. Again, this case is in the very early stages, but it is interesting to see the markets development supporting copyright infringement

VI. SUMMARY JUDGEMENT IN THOMSON REUTERS V. ROSS

A. Introduction

---

[522] Case 1:25-cv-01305, at Page 2 (2/13/20), https://admin.bakerlaw.com/wp-content/uploads/2025/02/ECF-1-Advance-Local-Media-Complaint.pdf.

[523] *Id* at 3.

This was the first filed AI case (in May 2020), and on February 11 2025, the court found copyright infringement and no fair use on summary judgement.[524] The opinion is written in the first person, as the Judge makes his way through the case and his decision. He begins, "A smart man knows when he is right; a wise man knows when he is wrong. Wisdom does not always find me, so I try to embrace it when it does—even if it comes late, as it did here."[525] Judge Bibas is referring to his early summary judgement opinion of 2023.[526] In that opinion Judge Bibas denied Thomson Reuter's motion for summary judgement on copyright and fair use, and moved the case towards trial.[527]

The case concerns Thomson Reuters' legal research platform, Westlaw. As part of its organization and added benefits, Westlaw has a complex headnote system that summarizes key point of case holdings. It includes a Key Number System, which is copyrightable, and since 1981, Thomson Reuters has been registering their headnotes since 1981.[528]

ROSS was a start-up that wanted to compete with Westlaw by using AI.[529] To train the system, ROSS needed a database of legal questions and answers. ROSS requested a license from Westlaw, which they refused.[530] So, ROSS went to LegalEase. [531] LegalEase crease "Bulk Memos" which are "lawyers' compilation of legal questions with good and bad answers."[532] LegalEase uses Westlaw's headnotes. LegalEase sold ROSS 25,000 Bulk Memos, which ROSS then used to train its AI search tool.[533] Thomson Reuters sued ROSS for copyright infringement.[534]

Judge Bibas frames the question: "whether the LegalEase Bulk Memo questions copied Thomson Reuter's headnotes or were instead taken from uncopyrightable judicial opinions."[535] The Judge did a comparison. Due to the fact that the questions and headnotes are sealed, he created an example based on *Feist Publication v. Rural Tel. Serv. Co.,* 499 U.S. 340,345 (1991).[536]

---

[524]Kate Knibbs, *Thomson Reuters Wins First Major AI Copyright Case in the US*, WIRED (Feb. 11, 2025), https://www.wired.com/story/thomson-reuters-ai-copyright-lawsuit/#:~:text=Thomson%20Reuters%20has%20won%20the,legal%20AI%20startup%20Ross%20Intelligence.

[525] *Thomson Reuters Enter. Ctr. GMBH v. ROSS Intelligence, Inc.*, No. 1:20-cv-613-SB, 2025 WL 458520, at *1 (D. Del. Feb. 11, 2025)

[526] *Thomson Reuters Enter. Ctr. GMBH v. ROSS Intelligence, Inc.,* 694 F. Supp. 3d 467 (D. Del. 2023), <u>vacated in part,</u> No. 1:20-CV-613-SB, 2025 WL 458520 (D. Del. Feb. 11, 2025)

[527] *Id*. at 481

[528] *Thomson Reuters*, 2025 WL 458520, at *2

[529] *Id*. at *7

[530] *Id*. at *1

[531] *Id.*

[532] *Id.*

[533] *Id.*

[534] *Id.*

[535] *Id*. at *2

[536] *Id.*

Question	West Headnote	Case Opinion
Does originality for copyright purposes mean that the work was independently created and has some minimal degree of creativity?	Originality, for copyright purposes, means that the work was independently created and has some minimal degree of creativity.	Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity.

B.   Direct Copyright Infringement

Judge Bibas held that ROSS infringed 2,243 headnotes.[537] Thomson Reuters owns a valid copyright in the headnotes. "Copyright registrations are 'prima facie evidence of the validity of the copyright' if made before or within five years after first publication of the work.'"[538] The court presumed then that Thomson Reuters had a valid compilation copyright. What is interesting about this argument is that you can have a copyright in the compilation, but the individual elements are not necessarily protected. But that is not addressed in the case. Why this becomes important is that Thomson Reuters is suing based on the individual headnotes and not the compilation. The factual dispute, according to the Judge, is whether the headnotes are still under copyright. Again, this strikes as odd. If the copyright registration covers the individual headnotes, there would be no way that the works would be in the public domain, as the term would be 95 years from publication. But we continue…

The Judge next turns to originality. Even though Thomson Reuters gets a presumption of validity because of the copyright registrations, could ROSS rebut the presumption that the headnotes did not meet the *Feist*'s minimal originality requirement? He finds the headnotes are original. He explains, "[a] headnote is a short, key point of law chiseled out of a lengthy judicial opinion. The text of judicial opinions is not copyrightable and even if it were, Thomson Reuters would not get that copyright because it did not write the opinions. But a headnote can introduce creativity by distilling, synthesizing, or explaining part of an opinion and thus be copyrightable. That is why I have changed my mind."[539] He explains that the headnotes are a compilation, and that Thomson Reuters have sufficient originality in the selection and arrangement to clear that law creativity bar.[540]

Moreover, each headnote has an individual copyright.

---

[537] *Id*. at *5
[538] *Id*. at *2
[539] *Id*. at *3.
[540] *Id*.

> That became clear to me once I analogized the lawyer's editorial judgement to that of a sculptor. A block of raw marble, like a judicial opinion, is not copyrightable. Yet a sculptor creates a sculpture by choosing what to cut away and what to leave in place. That sculpture is copyrightable. So too, even a headnote taken verbatim from an opinion is a carefully chosen fraction of the whole. Identifying which words matter and chiseling away the surrounding mass expresses the editor's idea about what the important point of law from opinion is. That editorial expression has enough 'creative spark' to be original. So, all headnotes, even any that quote judicial opinions verbatim, have original value as individual works.[541]

The judge does not grant summary judgement on any headnote that is verbatim copying. As to the Key Number System, that is original too.

Then the Judge turns to whether that were copying of original elements, looking first at actual copying and then at substantial similarity. He limits his analysis to 2,830 headnotes identified by Thomson Reuters, leaving 5,367 for trial. In the end, the Judge reviews 3,384 headnotes. ROSS actually used the copyrighted work to creates its own AI training data. The judge used an expert opinion report that the "Bulk Memo questions for this batch closely resemble the headnotes' text and the headnotes differ significantly from the text of the judicial opinions. Her findings suggests that these questions were created by copying Westlaw headnotes, not be summarizing the underlying opinions."[542] Moreover, the court compared how similar the Memo questions, headnotes and judicial opinions were, "one by one."[543] The Judge continues, "Having slogged through all 2,,830 headnotes, I grant summary judgement to Thomson Reuters on actual copying, finding actual copying of 2,243. Then he turns to substantial similarity, was original expression copied? The Judge explains, "As a lawyers and judge, I am myself an ordinary user of Westlaw headnotes. So, I am well positioned to determine substantial similarity here."[544] The Judge grants summary judgement for Thomson Reuters on 2,243 headnotes.

C. Fair Use

"None of ROSS's possible defenses hold water. I reject them all." So, begins Judge Bibas on the fair use analysis. Innocent infringement does not apply because there was a copyright notice on Westlaw's headnotes. Copyright misuse doesn't apply. The merge defense also does not apply. The scenes a faire defense does not apply either. That leaves fair use.

---

[541] *Id.*

[542] *Id.* at *11-12.

[543] *Id.* at *12

[544] *Id.* at * 13.

In his 2023 opinion, Judge Bibas denied summary judgement on fair use, but now he has changed his mind. For the first factor, He finds ROSS' use commercial and not transformative. Using *Warhol*, he finds that ROSS' does not have a "further purpose or different character" from Thomson Reuters. ROSS was using Westlaw headnotes to create a product that would compete with Westlaw.

The headnotes do not appear in the final product for consumers. The copying was part of an intermittent step. "ROSS turned the headnotes into numerical data about the relationships among legal words to feed into its AI."[545] Here's where it gets interesting. intermediate copying is permitted under fair use for computer programs. But the Judge does not see this as parallel. The computer cases are about computer code; this case is not. And the copying in the computer cases was "*necessary* for competitors to innovate."[546] The Judge cites *Google*, *Sony*, and *Sega*. So, instead of analyzing the fair use argument under the computer cases, the Judge analyzes it using *Warhol*.

Factors two and three are decided in ROSS' favor, as the works were factual in nature, and the final output doesn't include the Westlaw headnotes. Because of the *Author's Guild* case, you look to what is accessible to the public, and not the making of a copy.[547] Factor four is found for Thomson Reuters. The Judge defines the market as legal research platforms and the potential derivative mark as data to train legal AIs. But ROSS is competing with Westlaw to create a market substitute. In the end, the Judge rejects the fair use argument.

CONCLUSION?

ROSS' use was non generative – it doesn't create anything new. So, how much this opinion will impact on the generative AI cases is a question for another day. And we have many more cases working their way through the courts. Not to mention more licensing deals to seal. And if anyone is keeping score, here's some useful data.

As these generative AI cases progress through the litigation process, the original claims have developed based on the court's early findings. Each plaintiff originally asserted at least one cause of action concerning copyright infringement, violation of the Digital Millennium Copyright Act (DMCA), or some other related infraction (e.g., negligence). As part of the class exercise, we tried to chart what has happened so far.

---

[545] *Id.* at *17-18
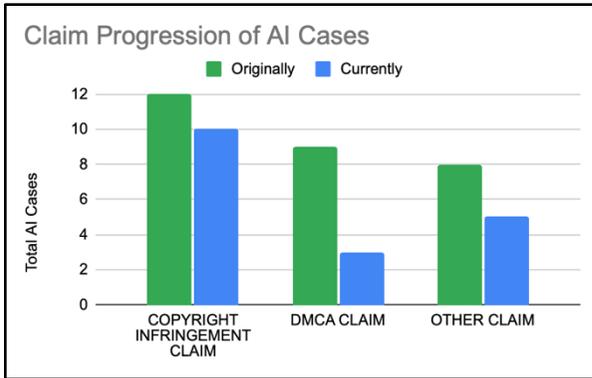[546] *Id.* at *18
[547] *Id.* at *21

Figure 1: Claim Progression of AI Cases

Figure 1 shows the total number of each type of original claim in comparison to the number of claims remaining as of February 2025. Twelve of the thirteen generative AI cases discussed here originally included a copyright infringement claim, and ten of those causes of action remain today. Other claims were slightly less successful, but five of the original eight remain. A more significant decline can be seen for the DMCA claims where only three of the original nine DMCA claims are still at issue. Figures 2, 3, and 4 further illustrate the varying success of each type of claim in remaining at issue in these cases today.
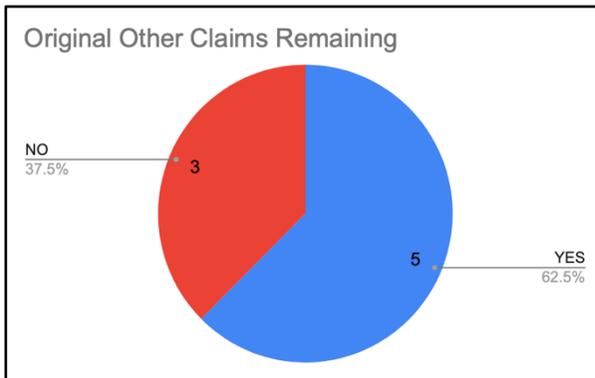

Figure 2: Original Copyright Infringement Claims Remaining

As seen in Figure B1, over 80% of copyright infringement claims survived initial rulings, while only 33% of DMCA claims remain. Other claims have lasted almost twice as often as DMCA claims, or roughly 62% of the time.

Figure 3: Original DMCA Claims Remaining

Figure 4: Original Other Claims Remaining

All of these generative AI cases are still in progress to some degree. While a majority of cases are still in progress independently, Figure 5 shows that five of the cases were combined with other similar matters, and one of the cases received partial summary judgment while awaiting a jury trial on other claims.
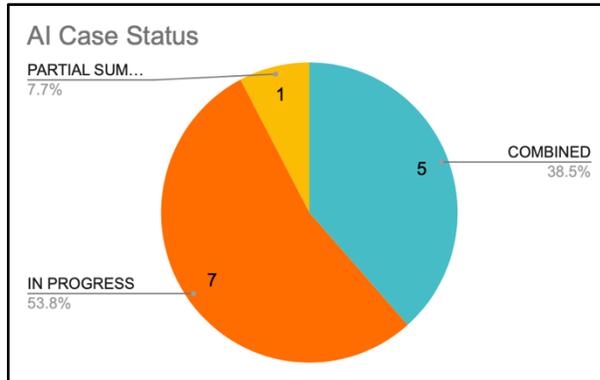
Figure 5: AI Case Status

In order to better understand the larger trends seen in these cases, data from Figures A and B can be divided into the categories of cases status shown in Figure 5. Figures 6, 7 and 8 compare the claim progression in cases with their current litigation status.
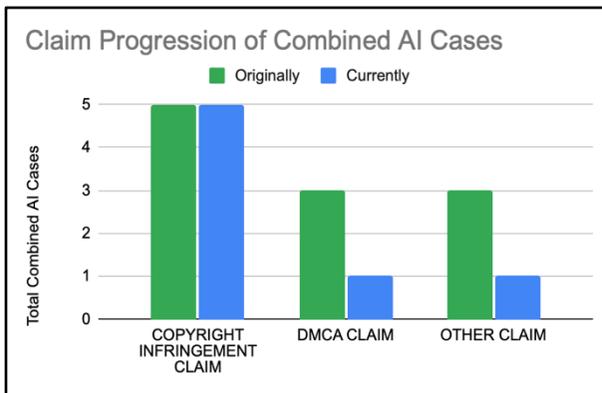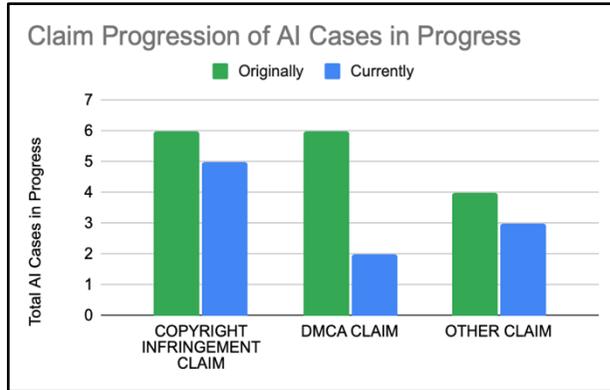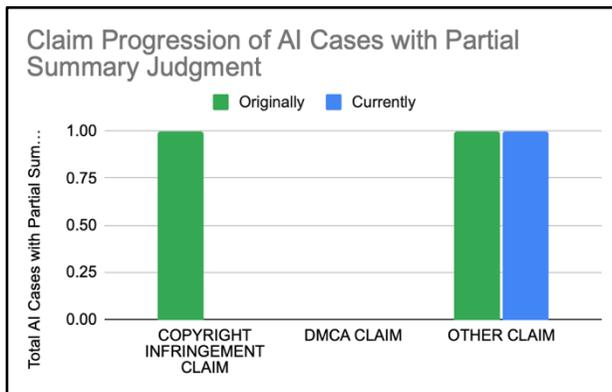

Figure 6: Claim Progression of Combined AI Cases

Figure 6 focuses on the generative AI cases that have been combined with other matters. Of the five cases that have been combined with others, all of them originally contained a copyright infringement claim and have maintained these claims throughout the process thus far. However, only one DMCA claim, and one other claim remains of the three of each that were originally filed, meaning one third of each of these types of claims remain in cases with a combined status.

Figure 7: Claim Progression of AI Cases in Progress

Figure 7 illustrates a similar trend for cases which are currently in progress independently and contain a copyright infringement or DMCA claim. Almost all copyright claims remain from the original filings, while only one third of DMCA claims remain. An important distinction between the data for combined cases and cases in progress independently is that those in progress have greater success maintaining other claims up to this point, as seen in Figure E where three out of four of claims remain.

Figure 8: Claim Progression of AL Cases
with Partial Summary Judgement

Figure 8 is less helpful in identifying larger trends in generative AI cases, as only one case has received partial summary judgment. Additionally, the complaint for this case originally did not contain a DMCA claim, so no data is available for that comparison.

***

Is there a conclusion? Not really, not yet. And so now we wait. The cases move through the court system. The third report is released by the Copyright Office. The Register of Copyright gets fired. And the semester ends.   Did the students get it right? What are they missing? Are we getting it right? What are we missin