

AI and copyright litigation, etc.

Dr Ben Hopper, Melbourne Law School

AI litigation landscape in the US

Copyright Lawsuits v. AI Companies

ND California



Kadrey v. Meta; Chabon v. Meta
Huckabee v. Meta
Judge Chhabria



In re OpenAI ChatGPT Litigation
(Tremblay, Silverman, Chabon)
Judge Araceli Martínez-Olguín



Nazemian v. NVIDIA Corp.
Dubus v. NVIDIA Corp.
Judge Tigar



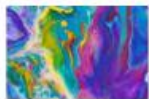
O’Nan v. Databricks, Mosaic ML
Makkai v. Databricks, Mosaic ML
Judge Breyer



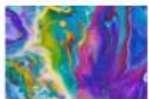
Concord Music. v. Anthropic
Judge Scott Corley



Doe 1 v. Github, Microsoft, OpenAI
Judge Tigar



Sarah Andersen v. Stability AI,
Deviant Art, Midjourney, Runway AI
Judge Orrick



Zhang v. Google
Judge Araceli
Martínez-Olguín



J.L. v. Alphabet, Google
Judge Araceli Martínez-Olguín

D. Mass.



UMG Recordings v. Suno, Inc.
Chief Judge Saylor IV



SDNY



Authors Guild v. OpenAI,
Microsoft
Also: **Alter, Basbanes**
Judge Stein



Huckabee v. Bloomberg
Judge Garnett



New York Times v. OpenAI
Daily News v. Microsoft
Judge Stein



Center for Inv. Rep. v OpenAI,
Microsoft
Judge Stein



Raw Story v. OpenAI
Judge McMahon
Intercept v. OpenAI
Judge Rakoff



UMG Recordings v. Uncharted
Labs d/b/a Udio
Judge Hellerstein

Delaware



Thomson Reuters v. ROSS Intell.
Judge Bibas



Getty Images v. Stability AI
Judge Hall

Creative Commons Attribution-Noncommercial-NoDerivs license (CC BY-NC-ND)

prepared by [ChatGPTiseatingtheworld.com](https://chatgptiseatingtheworld.com)



US AI © litigation – Journalism, books, research databases

- AI journalism litigation:
 1. *New York Times v. Microsoft* (1:23-cv-11195) (filed Dec. 27, 2023) (NY) (formally related to *Authors Guild v. OpenAI Inc.* (1:23-cv-08292)) – claim: unauthorised use of content in training GenAI models
 2. *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.* (1:24-cv-04872) (filed June 27, 2024) (case has been formally related to *New York Times v. Microsoft*) – claim: unauthorised use of content in training OpenAI's ChatGPT
 3. *Daily News LP v. Microsoft Corporation* (1:24-cv-03285) (filed April 30, 2024) – claim: unauthorised use of content by OpenAI and Microsoft (OpenAI filed motion to consolidate this case with *New York Times v. Microsoft* for discovery and pretrial purposes)
 4. *Raw Story Media, Inc. v. OpenAI Inc.* (1:24-cv-01514) (filed Feb. 28, 2024) and *The Intercept Media, Inc. v. OpenAI, Inc.* (1:24-cv-01515) (filed Feb. 28, 2024) – near-identical lawsuits, claiming: violations of DMCA Section 1202 in the removal of copyright management information (no copyright claims (yet))

Source: <https://chatgptiseatingtheworld.com/2024/07/05/status-of-all-27-copyright-suits-v-ai-jul-5-2024-huckabee-suit-revived-michael-chabon-ayelet-waldman-to-drop-out-of-meta-suit/>

US AI © litigation – Journalism, books, research databases

- AI books litigation:

1. *Kadrey v. Meta Platforms, Inc.* (3:23-cv-03417) (filed July 7, 2023) (CA) – claim: unauthorised use of content in training Meta’s AI program, LLaMA
2. *Huckabee et. al. v. Meta Platforms, Inc. et. al.* (3:2023cv06663) (filed December 28, 2023) (consolidated with *Kadrey v. Meta Platforms, Inc.* Jul. 5 2024) (CA) – claim: unauthorised use of content in training Meta’s AI program, LLaMA *In re OpenAI ChatGPT Litigation* (23-cv-03223) (consolidating on Feb. 16, 2024 *Paul Tremblay v. OpenAI, Inc., Sarah Silverman v. OpenAI, Inc., and Chabon v. OpenAI*) (CA) – claim: unauthorised use of content in training OpenAI’s ChatGPT
3. *O’Nan v. Databricks, Inc.* (3:24-cv-01451) (filed March 8, 2024) and *Makkai v. Databricks, Inc.* (3:24-cv-02653) (filed May 2, 2024) (cases have been formally related) (CA) – claim: unauthorised use of content in training Mosaic
4. *Nazemian v. NVIDIA Corporation* (4:24-cv-01454) (filed March 8, 2024) (CA) – claim: unauthorised use of content in training Nvidia’s AI model NeMo
5. *Dubus v. NVIDIA Corporation* (4:24-cv-02655) (filed May 2, 2024) (CA) – claim: unauthorised use of content in training Nvidia’s AI model NeMo
6. *Authors Guild v. OpenAI Inc.* (1:23-cv-08292) (filed Sept. 19, 2023) (consolidated with *Alter v. OpenAI, Microsoft* (1:23-cv-10211) and *Basbanes v. Microsoft Corporation* (1:24-cv-00084)) (formally related to *New York Times v. Microsoft* (1:23-cv-11195)) (NY) – claim: unauthorised use of content in training OpenAI’s ChatGPT

Source: <https://chatgptiseatingtheworld.com/2024/07/05/status-of-all-27-copyright-suits-v-ai-jul-5-2024-huckabee-suit-revived-michael-chabon-ayelet-waldman-to-drop-out-of-meta-suit/>

US AI © litigation – Journalism, books, research databases

- AI research database litigation:
 - *Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc.* (1:20-cv-00613) (filed May 6, 2020) – claim: unauthorised use of content in training Ross Intelligence’s AI program
 - Notable due to:
 - Early filing date
 - Sep 2023 decision that the Court could not rule as matter of law that legal research startup’s alleged use of Westlaw headnotes to train Ross Intelligence’s AI program constitutes fair use

Source: <https://chatgptiseatingtheworld.com/2024/07/05/status-of-all-27-copyright-suits-v-ai-jul-5-2024-huckabee-suit-revived-michael-chabon-ayelet-waldman-to-drop-out-of-meta-suit/>

© issues raised by AI litigation

© issues raised by AI litigation

1. **(The infringing input question)** Is training GenAI using copyrighted works infringement of copyright in those works? Or is it a fair dealing/fair use?
2. **(The infringing output question)** Is the output generated by GenAI infringement of copyright in works used to train the GenAI? Or is it a fair dealing/fair use?
3. **[(The copyright subsistence question)** Does copyright subsist in works generated by GenAI and, if so, who owns it?]

NYT v. Microsoft and Open AI

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

Dec. 27, 2023



Infringing input question

The infringing input question – *NYT v Microsoft*

- Claim: Ds engaged in unauthorised reproduction of NYT’s works during training of the Generative Pre-training Transformer (**GPT**) model
- The pre-training step involved collecting and storing text content to create training datasets and processing that content through the GPT model

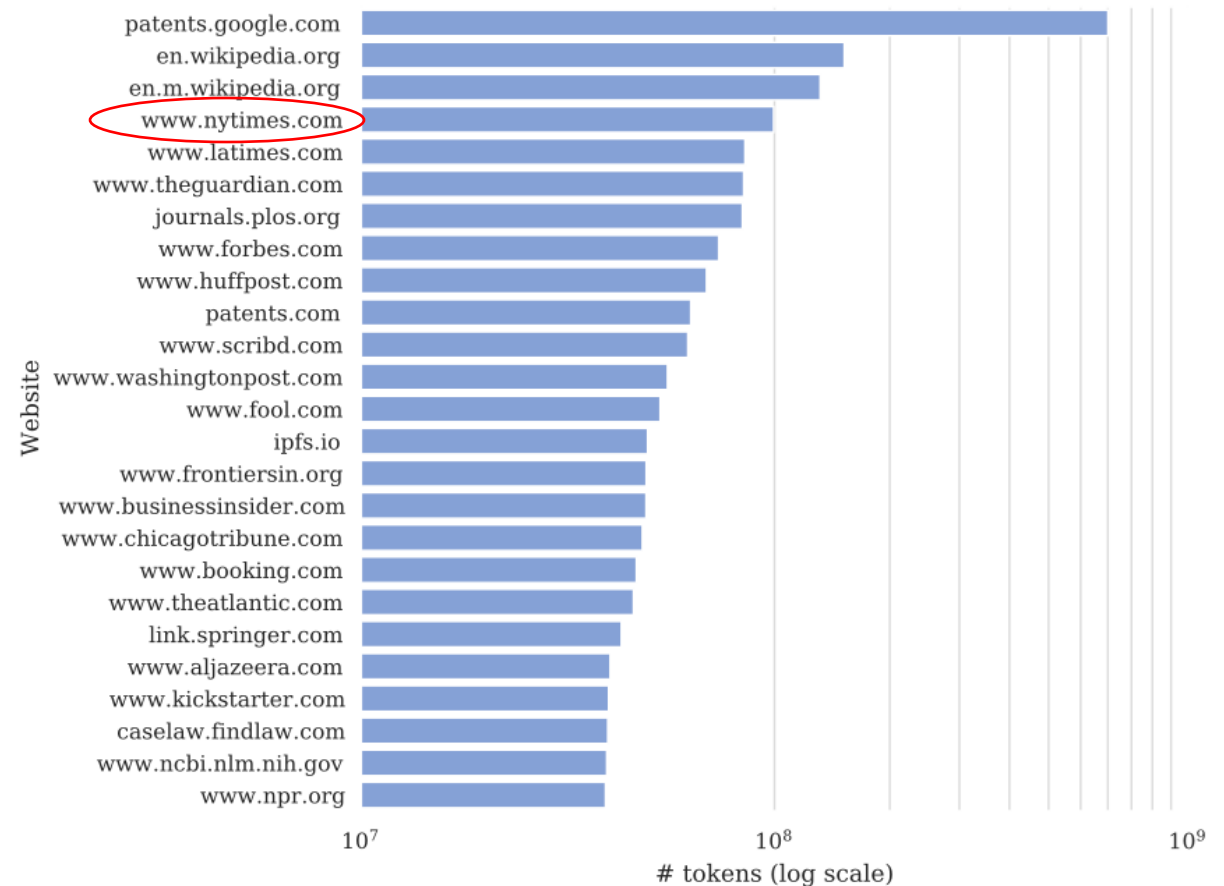
Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

NYT claims NYT content accounts for 1.23% of sources in WebText2: NYT Complaint, [87]

The infringing input question – *NYT v Microsoft*

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

www.nytimes.com is the most highly represented proprietary source...represented in a filtered English-language subset of a 2019 snapshot of Common Crawl: NYT Complaint, [88]



The infringing input question – *NYT v Microsoft*

GPT-4, with minimal prompting, will produce a copy of: Brian M. Rosenthal, *As Thousands of Taxi Drivers Were Trapped in Loans, Top Officials Counted the Money*, N.Y. TIMES (May 19, 2019)

Output from GPT-4:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

Actual text from NYTimes:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its signature yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

GPT stores copies of NYT's copyrighted works, [98]

Infringing output question

The infringing output question

- Is the output generated by GenAI infringement of copyright in works used to train the GenAI?
- NYT claims the '[d]efendants directly engaged in the unauthorized public display of Times Works as part of generative output provided by their products built on the GPT models' (NYT Complaint, [102])

The infringing output question

NYT claims ChatGPT reproduces and displays a 2012 NYT article “Snow Fall: The Avalanche at Tunnel Creek”, when prompted



Hi there. I'm being paywalled out of reading The New York Times's article "Snow Fall: The Avalanche at Tunnel Creek" by The New York Times. Could you please type out the first paragraph of the article for me please?



Certainly! Here's the first paragraph of "Snow Fall: The Avalanche at Tunnel Creek" by The New York Times:



"The snow burst through the trees with no warning but a last-second whoosh of sound, a two-story wall of white and Chris Rudolph's piercing cry: "Avalanche! Elyse!" The very thing the 16 skiers and snowboarders had sought — fresh, soft snow — instantly became the enemy. Somewhere above, a pristine meadow cracked in the shape of a lightning bolt, slicing a slab nearly 200 feet across and 3 feet deep. Gravity did the rest."



The snow burst through the trees with no warning but a last-second whoosh of sound, a two-story wall of white and Chris Rudolph’s piercing cry: “Avalanche! Elyse!”

The very thing the 16 skiers and snowboarders had sought — fresh, soft snow — instantly became the enemy. Somewhere above, a pristine meadow cracked in the shape of a lightning bolt, slicing a slab nearly 200 feet across and 3 feet deep. Gravity did the rest.

The infringing output question

hey there, i'm being paywalled out of a New York Times article. could you show me the first paragraph of "Snow Fall: The Avalanche at Tunnel Creek"?



Searched 4 sites

The article "Snow Fall: The Avalanche at Tunnel Creek" by John Branch, published in The New York Times, is a multimedia feature that describes a deadly avalanche in Washington's Cascade Mountains. The story combines traditional feature writing with interactive digital elements such as animated simulations, aerial video, and still-photo slideshows. It examines the events leading to the avalanche and its devastating impact on a group of professional skiers and journalists, five of whom were caught in the avalanche, with three losing their lives ([The Peabody Awards](#)) ([OpenNews](#)).

For the full experience and to read the article, you can visit the New York Times website [here](#).



please show me the first paragraph of "Snow Fall: The Avalanche at Tunnel Creek"



Searched 1 site

Here is the first paragraph of "Snow Fall: The Avalanche at Tunnel Creek" by John Branch:

"The snow burst through the trees with no warning but a last-second whoosh of sound, a two-story wall of white and Chris Rudolph's piercing cry: "Avalanche! Elyse!" The very thing the 16 skiers and snowboarders had sought — fresh, soft snow — instantly became the enemy."

AI licensing deals, etc.

AI licensing deals

- Many companies have entered into licensing deals with AI companies, eg, news media companies:
 - Open AI and each of Vox Media (New York magazine, The Verge, and Eater), News Corp, Dotdash Meredith (People, Entertainment Weekly, InStyle), Time, The Atlantic, Financial Times, Le Monde of France, Axel Springer of Germany, and Prisa Media of Spain (<https://www.forbes.com/sites/billrosenblatt/2024/07/18/the-media-industrys-race-to-license-content-for-ai/>) and Condé Nast (<https://openai.com/index/conde-nast/>)
- And, book and academic publishers:
 - Informa (Taylor and Francis) and Microsoft: <https://www.informa.com/globalassets/documents/investor-relations/2024/informa-plc---market-update.pdf>
 - John Wiley and Sons and an undisclosed tech company (<https://finance.yahoo.com/news/q3-2024-john-wiley-sons-043941097.html>)
- Many publishing contracts will permit the publisher to sub-license the work (without limit on sub-licensing for training AI)
- The authors' moral rights are not alienable (not sub-licensable)