

Reddit's Anthropic Lawsuit May Transform How AI Companies Use Your Data

Redefining AI training and testing the boundaries of fair use law

By Justin Smulison

IN A LAWSUIT THAT COULD RESHAPE THE RULES of AI training and fair use law, discussion and news aggregation website Reddit, Inc. filed a significant lawsuit in the Northern District of California on June 12 against AI startup Anthropic PBC, the company behind the Claude language model chatbot.

The lawsuit alleges that Anthropic unlawfully scraped and used vast amounts of Reddit's user-generated content—"millions, if not billions," according to the complaint—to train its AI models, without authorization, license, or compensation.

But Reddit did not file a copyright infringement claim to assert rights over its content. Reddit's complaint, however, accused Anthropic of systematically bypassing restrictions and refusing to engage in the standard commercial licensing arrangements agreed to by several other technology companies, such as Google. "Anthropic has copied and used, and continues to copy and use, Reddit's valuable and proprietary data without authorization or compensation," the complaint asserts.

The case moved to mediation in early August 2025. Its resolution could provide a clearer path for courts grappling with whether violating a platform's terms of service constitutes unlawful conduct, especially across state and international borders.

Reddit's lawsuit against Anthropic is more than just a dispute over scraping. It is a test case for the boundaries of fair use, contractual enforcement, and the business models that will define the generative AI era.

LEGAL FOUNDATIONS AND CORE CLAIMS

Roughly 116 million users comprised the global AI market in 2020, a *Statista Market Insights* survey noted. That figure nearly tripled to 314 million users by the end of 2024 and has shown no signs of relenting.

"AI is a popular topic for conversation, because it's paradigm-changing technology that's unfolding in front of us," said Jeremy

D. Bisdorf, chair of Taft's Technology and Artificial Intelligence industry group in Detroit and recognized by Best Lawyers® for Information Technology Law in Michigan. "Just as the Internet came around and it changed things, AI has the promise of doing the same thing. Anytime something new comes around that might affect how we live or work, people get really interested in it, and so do various companies with different goals."

One key objective is for tech companies to use AI to train their platforms and learning models. This included reading, repurposing and oftentimes redistributing information from copyrighted materials, like books and websites, which many authors argued was a violation of copyright law.

SCRAPING IN THE COURTS

Anthropic was founded in 2021 when former OpenAI employees decided they wanted to build their own AI company to develop safer, more ethical AI technology. As a public benefit corporation, they harnessed their AI models to do just that, and with partnerships with major companies like Amazon, it was poised for tremendous growth. According to *Bloomberg*, Anthropic was valued in late July at nearly \$170 billion, tripling from the \$61.5 billion mark it reached in the spring. Its financial hot streak was mirrored by various successful defenses in copyright law in 2025.

For example, in *Bartz v. Anthropic*, Judge William Alsup of the U.S. District Court for the Northern District of California found in June 2025 that "the purpose and character of using copyrighted works to train LLMs to generate new text was quintessentially transformative." Alsup stopped short of granting Anthropic a total victory when it came to piracy and books that were not purchased or legally obtained; *Bartz* will proceed to trial for the claim of pirated library copies (assuming no further motions defeat it).

Similarly, the defendant in *Kadrey v. Meta Platforms* (the parent company of Facebook and Instagram) withstood a two-year putative class action filed by 13 authors for including their books without permission in the training dataset for Meta's artificial intelligence

By framing their argument as protectors of third-party content, Reddit may be able to perpetuate itself as taking the ethical high ground.

'The amount of money they make on it is not as important as the moral and legal satisfaction of having Anthropic sign a license.'

product, LLaMA, based on public statements by Meta and others regarding the contents of the dataset.

On June 25, Northern District of California Judge Vince Chhabria signaled that Meta could continue its practice on these authors' works when he dismissed their claims for direct copyright infringement based on derivative work theory, vicarious copyright infringement and violation of Digital Millennium Copyright Act and other claims based on allegations that plaintiffs' books were used in training LLaMA.

TERMS OF SERVICE COMPLIANCE AND ENFORCEABILITY

To demonstrate that Anthropic had scraped Reddit content to train Claude, Reddit was likely paying close attention to how the courts were ruling in recent cases when it employed a new legal strategy—suing for breach of contract. Every major online platform (like Reddit and Meta) publishes Terms of Service (ToS) that typically restrict automated scraping, especially for commercial or AI-training purposes. The enforceability of these terms in court is a central legal issue: Can a website legally forbid scraping via its ToS?

"[Artificial intelligence] is the most cutting-

edge issue in copyright law in any of our lifetimes," said Jordan Greenberger, co-founder of Firestone Greenberger in New York. "The Reddit lawsuit is distinguishable from most of the other pending AI cases focused on copyright law. Copyright law can be used to address new technologies, but it's not always the clearest fit."

If an AI company bypasses stated protections (robots.txt, API limits), it risks ToS breach claims, which—while technically contract law, not copyright—can be powerful legal tools. Courts are now grappling with whether violating a platform's ToS truly constitutes unlawful conduct, especially across state and international borders.

"Here, Reddit is not making a copyright claim primarily because their terms of service are clear, that the user-generated content is not owned by Reddit," said Greenberger, who is also a member of the New York City Bar Association's Copyright & Literary Property Committee and the Copyright Society of the USA. "The authors of these various posts and conversations on the Reddit platform retain their rights and ownership interest in those works."

MONEY MATTERS AND REFRAMING REPUTATION

The content on Reddit is governed by a User Agreement and Privacy Policy, and its strategy to sue for its violation seems to coincide with its financial goals. Money talks when it comes to language models. In April 2023, Reddit announced its intentions to charge for its application programming interface (API), a previously free feature since 2008, causing a dispute as it targeted companies that crawl Reddit for data to train AI, especially large language models.

Womble Bond Dickinson Partner Ted Claypoole said that following Reddit's revenue stream—as well as other content providers—helps explain why they are so protective of their data.

In 2024, Reddit entered into a \$60 million deal with Google, which would allow the search engine to train its AI models on human posts; in 2025, it was estimated

that OpenAI paid \$70 million to Reddit for its licensing deal. Those two AI licensing deals comprised a reported 10% of Reddit's revenue, *AdWeek* reported.

"One of [Reddit's] arguments—and really their only major argument—is 'you're not allowed to do this because we have placed a value on the user comments,'" said Claypoole, who has been recognized by *The Best Lawyers in America* since 2017 for Technology Law. "And to prove that it's valuable to let people look at this, they can point to the multimillion-dollar licensing payments from other companies to do the same thing. So they will want Anthropic to do the same."

Beyond the commercial value, reputation matters as well. By framing their argument as protectors of third-party content, Reddit may be able to perpetuate itself as taking the ethical high ground.

"I think it makes it a bit more palatable for a judge to see Reddit as protecting their users rather than just what [they might] consider as commercial property," said Clark Hill Senior Attorney Chirag Patel. He noted that though Anthropic incorporated itself as a public benefit corporation, its behavior suggests otherwise, which can be leveraged by Reddit. "Now Reddit has taken the position that they're the defenders of their community, and that's, I think, the narrative that they want to advance."

REDDIT'S LONG-TERM STRATEGY

Reddit and Anthropic began mediation in early August, which could be advantageous for both companies, as well as other enterprises using and deploying AI. A trial could last for years, and various other legally ambiguous actions involving AI could be uncovered. A resolution reached in private, by contrast, would help set a precedent on regulating the scraping of user-generated content.

Claypoole believes that Reddit's overarching legal strategy was destined for mediation; if the company had lost at trial, Google, OpenAI and other companies would no longer have the incentive to pay millions for the protected

user information.

"Reddit likely wanted to force Anthropic into mediation, get them to sign a license, and then be able to say they got the big companies to adhere and keep it above board," said Claypoole, who also co-authored *The Law of Artificial Intelligence and Smart Machines*, which was published by the Business Law Section of the American Bar Association. He noted that a mediated resolution is likely the best-case scenario for Reddit and other gatekeepers and publishers of user content.

"The amount of money they make on it is not as important as the moral and legal satisfaction of having Anthropic sign a license," Claypoole said.

Considering the string of victories achieved by Anthropic, Meta and other companies accused of scraping, Claypoole expects all stakeholders—legal professionals, business owners, technologists, and users—will take note of developments in California's Ninth Circuit for guidance. The legal and economic lines drawn here and in similar cases will determine who controls the value of the internet's collective creativity and knowledge.

PUBLIC IMPACT AND THE ROAD AHEAD

For the public, the outcome of Reddit's mediated resolution will influence how user-generated content is treated not just legally, but ethically, as the source material feeding the next generation of AI systems. As AI becomes more embedded in daily life, the permissions, licenses, and protections negotiated will shape both individual privacy, public policy, and the broader digital economy.

"Many people don't want their personal information to be made just publicly available, and be able to be used by anybody without their consent," Bisdorf said. "And that idea is winning the day in many different state legislatures, which gets us back to the political point of whether it impedes the progress of the development of AI technology. If you're going to put a premium on privacy, your AI developments might lag behind a company or country who doesn't have the same philosophy on the privacy of data." ●