

Bartz v. Anthropic PBC

By Savannah Aleksic

Issue Statement

Everyone has a different stance on the topic of Artificial Intelligence (AI), especially as it continues to ebb its way into every facet of our online existence. AI uses a learning model designed to “learn from experience, identify patterns, and make decisions based on large volumes of data.”¹ Our topic of discussion today is Anthropic, an AI startup company founded in 2021 by several former employees of OpenAI. The case came about when a group of authors sued Anthropic in August 2024 in a class-action lawsuit, claiming that it illegally pirated copies of their books from Library Genesis (LibGen) and Pirate Library Mirror (PiLiMi) as tools to train its AI model.

Background

The AI company Anthropic has developed a large group of language models they refer to as Claude, which the company describes “aren't programmed directly by humans—instead, they're trained on large amounts of data.”² This is common practice for most AI systems. Think of an AI library collection as a kraken extending its tentacles in its reach to consume large amounts of information. In this case, the data was derived from copyrighted books, sourced from pirated sites.

The lawsuit was brought in the Northern District of California. The nonfiction authors Charles Graeber (*The Good Nurse*), Kirk Wallace Johnson (*The Feather Thief*), and thriller author Andrea Bartz (*We Were Never Here*) filed

1 <https://csuglobal.edu/blog/how-does-ai-actually-work>

2 <https://www.anthropic.com/research/tracing-thoughts-language-model>

the lawsuit against Anthropic. The copyrighted content included books that Anthropic purchased, scanned to create digital versions, and then destroyed. It is alleged that Anthropic also used over seven million digital copies of books acquired from pirating sites LibGen and PiLiMi to train its Claude LLMs. Plaintiffs sought damages and injunctive relief, as they claimed that Anthropic violated copyright law. In response, Anthropic maintained that their practices fell within the confines of fair use and, furthermore, were essential to the development of competitive AI technologies.

In June 2025, the court found two of Anthropic's accused activities were fair use: first, training Claude on plaintiffs' books was described as "exceedingly transformative", and second, digitizing purchased print books. The judge ruled "on summary judgment that using books without permission to train AI was fair use if they were acquired legally, but he denied Anthropic's request for summary judgment related to piracy—finding that the piracy was not fair use."³

In July, the court certified that all rightsholders of books acquired by Anthropic from LibGen and PiLiMi were registered with the U.S. Copyright Office and have ISBN or ASIN numbers. Since the books were pirated, this created a permanent, general-purpose library in their AI model that was not itself justifiable as being fair use. A trial was later scheduled to determine Anthropic's potential liability for piracy.

Judge Alsup found that any potential market harm would be no different than the harm caused by using the works for "training schoolchildren to write well," which could "result in an explosion of competing works." Judge Alsup stated that training "is not the kind of competitive or creative displacement that concerns the Copyright Act."

The court also concluded, "Anthropic had no entitlement to use pirated copies for its central library [the dataset used to train Claude]. Creating a permanent, general-purpose library was not itself a fair use excusing Anthropic's piracy."⁴

Since the books were registered with the U.S. Copyright Office in a timely manner, Judge Alsup determined that the millions of works downloaded were done so in an unlawful manner, and that Anthropic's obtainance of the works

3 <https://authorsguild.org/advocacy/artificial-intelligence/what-authors-need-to-know-about-the-anthropic-settlement/>

4 <https://www.bakerlaw.com/bartz-v-anthropic/>

from piracy sites may “erode the defense”, even in the context of using it to train an AI language model.⁵

In the settlement hearing on September 22nd, plaintiffs filed a supplemental brief that supported a motion for preliminary approval of a class settlement. This was in response to Anthropic’s agreement to pay a \$1.5B settlement earlier in the month. In the settlement, the sum amount for each work would amount to a gross recovery of \$3,000 per author. Due to the division of subsidiary rights, the authors will end up splitting 50/50 with their respective publishers, so authors will only take home around \$1,500 from the settlement. Judge Alsup expressed his discomfort with this settlement, as there were still important questions that had been left unanswered. People have a right to be concerned, as there has not been any proper legislation put in place to protect copyright holders when it comes to their work being used by a corporation to further advance their AI systems. We need to have a serious discussion and enact legislation to combat these issues to protect creators before more AI companies use immoral practices to their own advantage.

Argument/Opinion

While Anthropic’s AI training methods are challenging our conventional ideas of fair use, the court’s ruling has exposed an urgent ethical failure within our legal framework. This emphasizes an alarming precedent for the future of authorship in favor of innovation. One of the ethical considerations in using copyrighted material to train AI systems arises when the AI uses that information to create poetry, as part of Anthropic’s “AI biology” research findings. A concern would stem from the idea that a language model could consume a creative work, without permission, and use the artistic and linguistic inspiration from that work to create its own structures for publication rewards.

On the contrary, some may assert that AI systems trained using data aren’t much different than if you or I were to read Shakespeare and then write poetry using our newfound knowledge of poetic structure. However, AI training models are still just lines of code. AI may be able to replicate rhyme patterns and learn from literary works, but it does not have the one fundamental element essential in all forms of art: the human experience. AI artwork hinders the original copyrighted person’s creative liberties for monetary profit. If an

5 <https://www.insidetechnology.com/blog/2025/09/bartz-v-anthropic-settlement-reached-after-landmark-summary-judgment-and-class-certification>

AI system is allowed to use pirated creative works to further its own writing capabilities, is the AI work then considered the copyright material of the AI creator?

While many in the publishing community remain cautious on regular AI usage, whether for climate or moral reasons, there can be benefits in publishing fundamentals. As the ACSM (American College of Sports Medicine) describes, “Editors, editorial boards, publishers, and researchers will need to collaborate to ensure that AI tools are used responsibly, with appropriate safeguards in place to protect the authenticity of scientific research, uphold the quality of peer review, and ensure the fair treatment of all authors and reviewers.”⁶ AI is a strong and useful tool that can be utilized for more efficient and effective business practices, but it can also cause significant harm.

Conclusion

Just earlier this year, The United States District Court for Delaware rejected a fair use defense to the use of copyrighted works to train a natural language processing AI system in *Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc.*⁷ This should become the norm, as AI companies should be held to the same standards and ethical practices as everyone else, perhaps even more so.

It is no secret that the emergence of AI in publishing raises a number of concerns with transparency, accountability, and sustaining fairness in scholarship and creativity. At \$1.5 billion, *Bartz v. Anthropic* is now officially the largest U.S. copyright settlement in history. Not only is this an example of artists not being rewarded for their work financially, but the lack of accountability in AI companies provides an unhealthy environment going forward. Once we start diminishing meaningful authorship in favor of furthering technological advancement, we will only be pushing ourselves away from the authentic stories and experiences that make publishing so important.

6 <https://acsm.org/ai-ethics/>

7 <https://www.reedsmith.com/en/perspectives/2025/03/court-ai-fair-use-thomson-reuters-enterprise-gmbh-ross-intelligence>