

Publication

06/30/2025

Federal Court Finds That Training AI on Copyrighted Books is “Quintessentially” Transformative Fair Use

A federal district court in the Northern District of California has ruled that the use of lawfully acquired copyrighted works to train artificial intelligence (AI) large language models (LLMs) is a fair use under U.S. copyright law. The court’s opinion considered three separate uses to conclude that: (1) use of copyrighted material to train LLMs is a fair use; (2) conversion of purchased physical copies to digital copies to build a central library is fair use *where the original copies are destroyed and no copies are distributed*; and (3) download of pirated copies is not fair use, *regardless of whether the intended purpose was to eventually train LLMs*. As a caveat, the court repeatedly noted that the plaintiff authors made no allegations that any of the outputs of the LLM were infringing, and if they had, this “would be a different case.”

What is Fair Use?

Under U.S. copyright law, the copyright holder obtains the exclusive right to reproduce, distribute, perform, display, and prepare derivative works of the original work. But there are exceptions to exclusivity that are considered non-infringing “fair use.” Four factors are considered to determine fair use: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation

CLIENT SERVICES

Cybersecurity & Data Privacy

RELATED PEOPLE

David A. Wheeler

Alfred C. Tam

Kate H. Campbell

Josh Hanson

to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.

What Did Anthropic Do?

Anthropic is an AI company that provides a family of LLMs known as “Claude.” To train its LLMs, Anthropic set out to compile a central library of “all the books in the world.” Anthropic first knowingly downloaded pirated copies of books, and later purchased print copies, digitized them, and then destroyed the print copies. All of these copies were used to create a general-purpose research library, which Anthropic intended to use both for training LLMs and for other potential future uses. However, Anthropic did not distribute copies of the books, and Claude did not produce outputs that were exact copies or substantial knock-offs of the books.

How Did the Court Handle Each of Anthropic’s Separate Uses?

The court identified three uses as referenced above and considered each in turn.

First, on use for training LLMs, the court described this use as “exceedingly,” “quintessentially,” and “spectacularly” transformative and “among the most transformative many of us will see in our lifetimes.” The court emphasized that copyright law exists to “advance original works of authorship, not to protect authors against competition,” and analogized the training of LLMs to generate non-infringing outputs to the training of humans to be better writers. The author’s complaint on this point—so long as there was no actual infringing output—was, in the court’s view, “no different than it would be if they complained that training schoolchildren to write well would result in an explosion of competing works.”

Second, on Anthropic's conversion of purchased print copies to digital copies to store them in a central library, the court found fair use. The court emphasized that the conversion was merely a format change. The original copies were destroyed, and Anthropic did not distribute or share the digital copies publicly. In this sense, this use was essentially the same as keeping a library of purchased physical books, "but with more favorable storage and searchability properties."

Third, the court emphatically found no fair use for Anthropic's download of pirated copies to a central library. The court explained that "piracy of otherwise available copies is inherently, irredeemably infringing" and rejected Anthropic's argument that it intended to eventually use those copies in a transformative manner to train LLMs. As the court put it, "Anthropic is wrong to suppose that so long as you create an exciting end product, every 'back-end step, invisible to the public,' is excused."

Copyright Office's Report on Copyright and Artificial Intelligence

This decision comes on the heels of the U.S. Copyright Office's Report on Copyright and Artificial Intelligence.[1] Part 3 of that report underscores the necessity of balancing innovation in AI with the protection of creators' rights. The report highlights the importance of developing legal and market-based solutions to address the challenges posed by the use of copyrighted materials in AI training. Key policy considerations include:

- Fair use limitations in commercial contexts where AI outputs could substitute original works.
- Market impacts of large-scale scraping of copyrighted content for AI training and its adverse effects on the market for original works.

- Dataset transparency and licensing frameworks to facilitate lawful use of copyrighted materials, ensuring that creators are compensated and their rights respected.
- Clearer legal guidelines to delineate the boundaries of fair use in the context of AI training, helping both creators and developers navigate the legal landscape.

As AI continues to proliferate and as courts continue to grapple with how and where to draw the line between AI innovation and creators' rights, the legal landscape will remain uncertain. Should you have any questions regarding the intersection of U.S. copyright law and AI, its impacts on your business, or any compliance obligations it creates, please reach out to David A. Wheeler, Alfred C. Tam, Kate H. Campbell, Josh A. Hanson or your Neal Gerber Eisenberg attorney.

[1] <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>

The content above is based on information current at the time of its publication and may not reflect the most recent developments or guidance. Neal, Gerber & Eisenberg LLP provides this content for general informational purposes only. It does not constitute legal advice, and does not create an attorney-client relationship. You should seek advice from professional advisers with respect to your particular circumstances.