

June 26, 2025

Fair Use in Training AI Models

First Decisions by U.S. Courts Addressing Whether Using Copyrighted Materials to Train AI Models Is Fair Use

SUMMARY

On June 23, 2025, the U.S. District Court for the Northern District of California (Judge Alsup) held in *Bartz v. Anthropic*¹ that Anthropic's use of copyrighted books to train its generative artificial intelligence ("AI") models was protected by fair use under U.S. copyright law. In particular, Judge Alsup emphasized that Anthropic's use was "spectacularly" and "exceedingly" transformative. Two days later, on June 25, a different judge of the same court (Judge Chhabria) held in *Kadrey v. Meta Platforms*² that Meta's use of copyrighted books to train its models was likewise protected by fair use, finding that there was "no serious question" that Meta's use of copyrighted materials was "highly transformative."

These are the first decisions by U.S. federal courts to address whether a central feature of many AI companies' business models—the use of copyrighted materials to train generative AI models—infringes the copyrights in such materials. Whether these opinions survive on appeal or what parts, if any, of their reasoning are followed by other courts facing similar questions remains to be seen. These early cases, however, represent a positive development for the AI industry seeking clarity on these significant questions.

BARTZ v. ANTHROPIC & KADREY v. META

A. BACKGROUND

1. *Bartz v. Anthropic*

Anthropic, a company that develops frontier AI models, allegedly downloaded millions of copyrighted books (including copies obtained via the Books3, LibGen, and PiLiMi libraries) and purchased millions more hard-copy books, which it scanned. These copies were pooled into a central "research library" intended to be retained "forever." Anthropic engineers allegedly used various subsets of this "central library" to train the large language models ("LLMs") underpinning Claude, the company's flagship generative AI model. The

plaintiffs, Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, are authors whose books are included in Anthropic's central library and were used to train Anthropic's LLMs. The plaintiffs filed a putative class action, alleging that Anthropic infringed on their copyrights by (1) training its LLMs on the plaintiffs' copyrighted materials, (2) using unauthorized copies of their books to create a central library of "potentially useful content," and (3) scanning and digitizing purchased hard copies of their works to add to the central library.

Claude includes filtering software to prevent the model from generating outputs that are exact copies or substantially similar to its training data. The allegations in the case were, therefore, limited to copyright infringement by training the LLMs. No allegations were made that Claude's outputs were also infringing. Anthropic argued, in its motion for summary judgment, that its use of the plaintiffs' books was fair use under U.S. copyright law.

2. *Kadrey v. Meta Platforms*

Meta is a social media company that in 2023 began developing LLMs called Llama. Similar to the plaintiffs in *Bartz*, the plaintiffs here are authors who allege that Meta infringed their copyrights by using their works to train Llama. They filed a putative class action, advancing a variety of claims, including a claim for direct copyright infringement. Plaintiffs moved for summary judgment on their copyright infringement claim, and Meta cross-moved for summary judgment on its fair use defense.

Similar to Claude, Llama contains filtering functionality intended to prevent it from providing copyrighted outputs to its users.

B. OPINIONS

Fair use is an affirmative defense to copyright infringement under U.S. law and is codified in the U.S. Copyright Act. The fair use doctrine requires courts to balance four factors: (1) the purpose and character of the use; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used; and (4) the effect of the use upon the market for or value of the copyrighted work.³ The Supreme Court has routinely emphasized that "fair use is a flexible concept, and its application may well vary depending on context."⁴

Judge Alsup accepted Anthropic's fair use defense as to (i) training Anthropic's LLMs on the plaintiffs' copyrighted materials and (ii) scanning purchased copies of the plaintiffs' works to add to the central library. But he allowed the plaintiffs' claim regarding use of copies of their books obtained from the third-party libraries to proceed to trial. Judge Chhabria similarly accepted Meta's fair use defense as to Llama's training on the plaintiffs' copyrighted materials.

Both judges found that the second factor (nature of the work) universally cut in the plaintiffs' favor because the expressive nature of their work brought this material "closer to the core of intended copyright protection." But the crux of their analyses turned on the first, third, and fourth factors of the fair use test.

1. The Copies Used to Train Specific LLMs

With respect to the first factor (purpose and character of use), Judge Alsup found that Anthropic's use to "map statistical relationships between every text-fragment . . . and sequence of text-fragments" was "spectacularly" transformative. He reasoned that Claude did not replicate the plaintiffs' works; instead, Claude produced new text in response to user prompts "as if it were a human." He noted that the plaintiffs could not identify a single instance in which Claude's outputs had infringed a specific copyright. Moreover, Judge Alsup favorably likened Claude's learning process to that of a human who internalizes a book's content and later draws on it for future inspiration without directly copying it.

On the first factor, Judge Chhabria similarly found that Meta's use of copyrighted materials to train Llama was "highly transformative."

On the third factor (amount used), Judge Alsup held that Anthropic's copying of the books was "reasonable in relation to the purpose of the copying" and "compelling" in light of the large volume of text that Anthropic needed to train its LLMs. He cited the decision from the U.S. Court of Appeals for the Second Circuit in *Authors Guild v. Google Books*⁵ for support, noting that Anthropic, like Google in that case, only copied the entirety of the materials for internal use and did not make it available to consumers. Judge Chhabria came to the same conclusion on the third factor.

The two judges diverged most sharply over the fourth factor (effect on market). Judge Alsup rejected the plaintiffs' arguments that Claude's existence or outputs would compete with the plaintiffs' books or significantly impact their potential market. He held that the plaintiffs' argument was "no different" from an argument that "training schoolchildren to write well would result in an explosion of competing works." He concluded that such competition was not the type of "displacement that concerns the Copyright Act."

By contrast, Judge Chhabria held that "indirect substitution is still substitution." He found that Llama's potential to create works similar to plaintiffs' that could dilute the market for plaintiffs' works could weigh against a finding of fair use under the fourth factor (effect on market). But Judge Chhabria did not rule on the fourth factor because he found the plaintiffs had neither made the indirect substitution argument nor provided evidence to support it.

On balance, both judges held that copying the plaintiffs' works to train Anthropic's LLMs was fair use.

2. The Unauthorized Copies Used to Build a Central Library

Judge Alsup also held that Anthropic's use of unauthorized copies of copyrighted materials to build a central library was not fair use. He held that there was nothing transformative about "[p]irating copies to build a research library without paying for it." He further added that virtually any use of reproduced copies would, by definition, impermissibly replicate the original work, because that material could just as easily have been purchased and copied to fulfill Anthropic's stated purpose of building a central library. Judge Alsup also held that Anthropic's downloading of reproduced copyrighted works "plainly" usurped the market for these

works because Anthropic took for free what it should have had to purchase. Accordingly, all four factors pointed against fair use.

There is a lack of clarity in *Bartz* on whether training AI models on unauthorized copies obtained from third-party libraries would be fair use, as Judge Alsup did not need to reach this question.

3. The Purchased Copies Converted from Print to Digital to Build a Central Library

Judge Alsup held that Anthropic's conversion of purchased books from print to digital was fair use. This "mere format change" was considered transformative because this digitization saved space and enabled searchability. He found that copying of the entire work was justifiable to create a searchable library; if the entire book was not digitized, a searchable library would not be possible. He held that the fourth factor (effect on market) was neutral because there was insufficient evidence that Anthropic intended to distribute the digital copies in a way that would impede plaintiffs' ability to sell additional copies of their works. Overall, Judge Alsup held that the format conversion was fair use.

C. IMPLICATIONS

Bartz and *Kadrey* are the first decisions by U.S. federal courts to rule on fair use in the generative AI context. The outcomes are positive developments for AI companies, as they are the first decisions to hold that the use of copyrighted materials to train AI models can, at least in some circumstances, qualify as fair use under U.S. copyright law.

But these decisions should not be overread for four reasons. First, the judges' reasoning is tied to specific characteristics of the models at issue in their cases, and both judges made clear that their holdings would not apply in cases where the AI model replicated copyrighted material as outputs to users. Second, *Bartz* leaves open questions about the sourcing of training materials, as Judge Alsup's opinion is unclear on whether training AI models on unauthorized copies obtained from third-party libraries would be fair use. Third, both *Bartz* and *Kadrey* are district court rulings; they could be reversed on appeal, and other courts—including eventually the U.S. Supreme Court—may reach different conclusions. Finally, Judge Chhabria indicated that he may have ruled for plaintiffs on the fourth factor (market effect) if the plaintiffs had raised different arguments about, and provided sufficient evidence to support, how Llama threatened to dilute the market for their works. Judge Chhabria's statements on the fourth factor to this effect may leave the door open for potential claims alleging dilution or usurpation of content creators' markets.

In short, both cases represent only an initial step in clarifying whether the use of copyrighted materials in training AI models is protected by fair use under U.S. law.

* * *

ENDNOTES

- ¹ 3:24-cv-05417 (N.D. Cal. filed Aug. 19, 2024).
- ² 3:23-cv-03417 (N.D. Cal. filed July 7, 2023).
- ³ 17 U.S.C. § 107.
- ⁴ *Andy Warhol Found. for the Visual Arts v. Goldsmith*, 598 U.S. 508, 527 (2023) (internal citation omitted).
- ⁵ *Authors Guild v. Google Books*, 804 F.3d 202 (2d Cir. 2015).

SULLIVAN & CROMWELL LLP

ABOUT SULLIVAN & CROMWELL LLP

Sullivan & Cromwell LLP is a global law firm that advises on major domestic and cross-border M&A, finance, corporate and real estate transactions, significant litigation and corporate investigations, and complex restructuring, regulatory, tax and estate planning matters. Founded in 1879, Sullivan & Cromwell LLP has more than 900 lawyers on four continents, with four offices in the United States, including its headquarters in New York, four offices in Europe, two in Australia and three in Asia.

CONTACTING SULLIVAN & CROMWELL LLP

This publication is provided by Sullivan & Cromwell LLP as a service to clients and colleagues. The information contained in this publication should not be construed as legal advice. Questions regarding the matters discussed in this publication may be directed to any of our lawyers or to any Sullivan & Cromwell LLP lawyer with whom you have consulted in the past on similar matters. If you have not received this publication directly from us, you may obtain a copy of any past or future publications by sending an e-mail to SCPublications@sullcrom.com.