

July 1, 2025

Two Federal Courts Rule That Reproduction of Books to Train LLMs Is Fair Use — But with Caveats and Strikingly Different Views

By: [Jedediah Wakefield](#), [David L. Hayes](#), [David A. Bell](#),
[Charles Moulins](#)

What You Need To Know

- In first-of-their-kind rulings, two California federal judges concluded that reproducing copyright-protected materials to train large language models was fair use. Both judges emphasized the transformative nature of the training process and held that the plaintiffs had failed to show any harm to the market for their books.
- Both judges held that copyright holders cannot claim harm to a potential market to license their

Related Professionals



[Jedediah Wakefield](#)
Partner · Litigation



[David L. Hayes](#)
Partner · Intellectual Property



[David A. Bell](#)
Partner · Corporate

works for training purposes as a basis for defeating fair use.

- The courts left the door open to liability from activity related to model training depending on the facts, but for very different—and conflicting—reasons.
- The judges parted ways on two issues of first impression: (1) whether the “dilution” of the market for copyrighted works from inexpensive, AI-generated content is relevant to the fair use analysis, and (2) whether initially acquiring and storing copyrighted materials from so-called “pirate” sources defeats fair use.

Related Practices

- [Litigation](#)
- [Complex Litigation](#)

Related Industries

- [AI & Machine Learning](#)

[🔗 Share](#)

→ [Subscribe](#)

In separate high-profile actions brought by authors against Anthropic and Meta, two California federal judges ruled that the reproduction of copyright-protected books to train large language models (LLMs) was fair use that did **not** give rise to any copyright infringement. But, for very different reasons, both judges left the door open to possible copyright infringement liability in future proceedings.

In [Bartz v. Anthropic PBC](#), the court allowed infringement claims premised on Anthropic’s creation of a “central library” of so-called “pirated” copies of books to proceed to trial, while the court in [Kadrey v. Meta Platforms, Inc.](#), rejected claims that initial acquisition of allegedly “pirated” books defeated Meta’s fair use defense. Meanwhile, the *Anthropic* court rejected the

idea that competition from *non-infringing* AI-generated content would create cognizable market harm under the fair use analysis, while the *Meta* court reached the opposite conclusion. Although the *Meta* court concluded plaintiffs had not shown such harm in that case, the court provided a roadmap for other authors to pursue that novel theory in other lawsuits.

Background

In *Anthropic* and *Meta*, the plaintiffs are authors who allege that their books were included without authorization in datasets used to train Anthropic's Claude and Meta's Llama LLMs. Anthropic and Meta moved for summary judgment (disposition before trial based on undisputed facts), arguing that any reproduction of the plaintiffs' works to train LLMs was "fair use" under § 107 of the Copyright Act, and thus did not give rise to any liability for copyright infringement.

Both courts agreed that LLM training is transformative, and that the authors failed to show any market harm.

The *Anthropic* and *Meta* judges determined that Anthropic and Meta's reproduction of books for LLM training was fair use. To reach that conclusion, they weighed four statutory factors spelled out in § 107:

1. The purpose and character of the use
2. The nature of the copyrighted work
3. How much of the copyrighted work is used
4. The effect of the use upon the potential market for or value of the copyrighted work

Both judges focused on factors 1 and 4, which are decisive in many cases.

The judges agreed that factor 1 (purpose of the use) weighed heavily in favor of fair use, as *Meta* and Anthropic's reproduction of the books served a completely different purpose than the books themselves. The *Meta* court explained that "[t]he purpose of Meta's copying was to train its LLMs, which are innovative tools that can be used to generate diverse text and perform a wide range of functions," whereas the purpose of the books "is to be read for entertainment or education." The *Anthropic* court similarly held that developing LLMs is "spectacularly" transformative, and that "[t]he technology at issue was among the most transformative many of us will see in our lifetimes."

Both courts held that factor 4 (effect on the market) also favored fair use, because—on the record in those cases—training the LLMs did not harm the market for the authors' works. The courts noted that neither the Claude nor the Llama models could generate free copies of the books. The courts also rejected the plaintiffs' argument that using their books without permission harmed them because they could otherwise have requested licensing fees for that use. The *Meta* court explained that claiming harm to a market to license the very use that was at issue was "circular" and would improperly favor copyright owners in every fair use case.

Each court held that certain conduct may give rise to copyright claims, while disagreeing on key issues.

Both courts left the door open to copyright infringement liability for training LLMs, but they diverged on two important issues that have broad implications for model training: (1) whether potential “dilution” in the market for copyrighted works from easily-produced, AI-generated content is a cognizable harm under the fourth fair use factor (effect on the market), and (2) whether the initial acquisition of copyrighted materials from so called “pirate” sources defeats fair use.

Market Dilution: With no evidence that Anthropic’s LLMs generated infringing output, the *Anthropic* court rejected the idea that competition from a proliferation of **non-infringing** books created from LLMs is a cognizable harm under copyright law (irrespective of any broader societal policy implications). That result was consistent with Ninth Circuit fair use authority. See, e.g., *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1523 (9th Cir. 1992) (finding fair use from copying video games for compatibility purposes, despite the ultimate creation of competing video games: “[i]t is precisely this growth in creative expression, based on the dissemination of other creative works and the unprotected ideas contained in those works, that the Copyright Act was intended to promote.”). Rejecting arguments that competition from a flood of non-infringing LLM-generated books would create cognizable harm to the market for the original works, the *Anthropic* court held that the “[a]uthors’ complaint is no different than it would be if they complained that training schoolchildren to

write well would result in an explosion of competing works.”

The *Meta* court fundamentally disagreed, concluding that the true harm from LLMs trained on copyrighted materials is “market dilution,” where the LLMs “enable the rapid generation of countless works that compete with the originals, even if those works aren’t themselves infringing.” Although there was no evidence of that harm in the record in the *Meta* case, the judge noted that other authors may have better luck building that evidentiary record, concluding that the authors and their counsel “made the wrong arguments and failed to develop a record in support of the right one.”

Initial Acquisition: The *Anthropic* court held that it was fair use for Anthropic to buy physical books, scan them and destroy the physical copies, and then use the scanned copies for LLM training. But when Anthropic initially acquired books by downloading them from unauthorized sources—what the court called “pirating”—it was **not** fair use. For that reason, the *Anthropic* court allowed infringement claims premised on Anthropic’s creation of a “general-purpose” “central library” of “pirated” books to proceed to trial. The judge also suggested, without deciding, that downloading unauthorized copies of books may be “inherently, irredeemably infringing even if the pirated copies are immediately used for the transformative use [i.e., to train LLMs] and immediately discarded.” But the *Meta* court disagreed that unauthorized initial acquisition of books could independently give rise to liability for copyright

infringement, because “that downloading must still be considered in light of its ultimate, highly transformative purpose: training [LLMs].” That court granted summary judgment that Meta engaged in fair use by using the downloaded material to train its models, despite plaintiffs’ claims that Meta knowingly downloaded free books from so-called “pirate” sites.

Thus, despite a partial win for Anthropic and complete win for Meta, both courts cautioned—albeit for fundamentally different reasons—that LLM training may give rise to liability for copyright infringement under certain circumstances.

Next Steps

Appeals are likely given the *Meta* and *Anthropic* courts’ split on core copyright issues related to so-called “pirate” copying for LLM training and “market dilution” from LLMs potentially creating non-infringing but competing works. Courts in other copyright cases currently pending against AI developers will need to address these issues, increasing the importance of appellate court guidance.

Related Insights

Publications

DeepSeek, Model Distillation, and
the Future of AI IP Protection

February 03, 2025

Publications

U.S. Imposes New Export Controls
on Advanced Artificial Intelligence
Technologies and Chips

February 05, 2025



[View more related insights](#)