

Publications

July 11, 2025 • Updates

AI vs. Authors: Two California Judges, Two Directions and More Uncertainty on Fair Use and Copyright

Key Takeaways

- **Courts Lean Toward Fair Use for AI Training:** Two California rulings suggest that using copyrighted works to train artificial intelligence (AI) may be considered fair use if outputs are transformative and do not replicate the original content.
- **Pirated Libraries Raise Legal Risks:** While courts accepted some use of pirated works for transformative AI purposes, they strongly criticized maintaining central libraries and training AI with pirated content for non-transformative purposes, signaling potential legal vulnerability.
- **Legal Uncertainty Remains:** With no clear precedent or updated legislation, both authors and tech companies face ongoing uncertainty; future guidance will likely need to come from Congress or the Supreme Court.

The rapid advancement of AI large language models (LLMs) depends heavily on ingesting vast amounts of textual data, often scraped from books, websites and online libraries. Two recent rulings from the United States District Court for the Northern District of California delivered diverging views on one of the most disputed issues in AI technology: can tech companies use copyrighted books to train AI models without permission or a license from the authors of those books?

The two cases, *Bartz v. Anthropic PBC* and *Kadrey v. Meta Platforms, Inc.*, involved authors suing AI tech companies who were training their LLMs by sourcing high volumes of copyrighted works from either pirated or shadow libraries. At the center of the courts' fair use considerations was the "transformative use" test in analyzing the defendants' fair use defense. The transformative use test considers whether a secondary work significantly alters the original copyrighted work by adding something new, such as a new meaning, expression, or message, as opposed to merely copying the work. Essentially, if a copyrighted work was repurposed in a way that adds new meaning or function without affecting the market for the original work, then the use may be considered fair use and not a violation of the Copyright Act.

In regard to ingesting and using a high volume of copyright works (in the millions) to train AI models, both courts appeared more forgiving of such training when the results include

Related People

- Melenie Van
- Jay E. Heidrick

Related Capabilities

- Intellectual Property
- Intellectual Property Litigation
- Artificial Intelligence & Machine Learning

(1) the generation of textual outputs that did not infringe or substitute for the original copyrighted works and (2) the development of innovative tools such as language translators, email editors and creative assistants. Both courts found such use — even if the copyrighted works were sourced from pirated or shadow libraries — to be transformative, constituting fair use.

However, both courts were also concerned about the creation of a library and training AI using pirated works. The *Bartz* court criticized Anthropic for its act of retaining a subset of pirated books in its central library despite concluding that those sets of works would never be used for training LLMs, and found such conduct did not constitute fair use. The *Kadrey* court strongly suggested in *dicta* that, where training LLMs does not constitute fair use, AI developers will still figure out a way to license copyrighted works — especially given how beneficial such training is, as emphasized by Meta.

At least these two decisions show that courts are willing to consider LLM training fair use under certain conditions. Both decisions focused on the transformative output of the LLM that included extensive steps such as statistical analysis, text mapping and computer analysis beyond merely reproducing the work. Both decisions also noted the limited ability of the LLMs to produce more than snippets of the work in their results. However, both also focused on the pirated nature of the libraries and AI training and held (or indicated) that such conduct done for non-transformative purposes would not be fair use. The lack of unified precedent or legislative guidance means future decisions in other jurisdictions may not be uniform until the issue is resolved at the appellate level, likely by the Supreme Court. Congress will likely face renewed pressure by both creators and AI tech companies to update copyright law to address AI as the current Copyright Act is not equipped to do so.

Until there is guiding precedent or legislative action, tech companies and copyright holders alike are watching closely. Authors should ensure their works have copyright protection. For tech companies, the consequence of infringement is high as the statutory damages provided by the Copyright Act can be significant. To avoid this exposure, costly litigation and reputational risk, AI companies should focus both on the manner in which they create libraries and the LLM's outputs. To lessen scrutiny, LLM libraries should not be created from unlicensed or unprotected works as both decisions highly criticized creating large libraries from pirated works for non-transformative uses. However, outputs based on text mapping, statistical analysis and other computer functions that do not reproduce significant portions of the work appear more likely to constitute fair use. Tech companies should consider whether any licenses for library creation will also cover LLM outputs. But given the vast amount of information needed to create an LLM library, this may be impractical or cost prohibitive. To cut down on logistics and licensing costs, AI companies should work with already established databases and libraries when sourcing materials to train their LLMs.

In the end, these decisions highlight the old adage that the only certainty about litigation is uncertainty. Until we have guiding precedent or legislative action, both authors and AI companies remain uncertain as to the scope of their rights regarding the use of works in the AI community.