



ALERT · JUNE 25, 2025

District Court Issues AI Fair Use Decision: Using Copyrighted Works to Train AI Models Is Fair Use, but Using “Pirated” Copies to Build a Central Library Is Not

BY Robert D. Carroll Stefan Mentzer Timothy Keegan Eunbi Choi

A federal district court in San Francisco ruled that training AI models with copyright-protected works is fair use.¹ On June 23, 2025, Judge William Alsup ruled that Anthropic did not infringe the books of three authors used to train its AI models; the court found the training “exceedingly transformative” and that the fair use factors as a whole weighed for Anthropic. The court also ruled that Anthropic’s use of “pirated” copies to build a digital central library of works was not fair use, and ordered the question of those uses to proceed to trial.

Judge Alsup’s decision has important implications for the training of AI models. It is a win for the argument that training AI models is fair use. But Judge Alsup hinted that there may be circumstances where training LLMs is not fair use: in *dicta*, he suggested that using “pirated” copies to train LLMs is not fair. The decision also suggests the copying of works for purposes other than for training LLMs may give rise to liability. The decision does not address whether the “outputs” of AI products like Anthropic’s Claude are infringing or fair use.

Background

Anthropic offers the widely-used AI service Claude. Anthropic trained large language models underlying various versions of Claude on millions of copyright-protected books. Anthropic sourced these books by:

- Obtaining 7 million unauthorized copies, which the court repeatedly referred to as “pirated.”
- Bulk-purchasing millions of lawful copies of print books, which Anthropic then scanned to digital and discarded.

Anthropic used these copies to create a “central library.” The company kept the library for general purposes, including – but not limited to – training LLMs.

Three authors filed a putative class action complaint against Anthropic for copyright infringement. The case was assigned to Judge Alsup, who is experienced in copyright matters. Judge Alsup presided as district judge in another notable fair use case, *Oracle v. Google*.

Anthropic moved for summary judgment on the issue of fair use, before class certification. Judge Alsup issued his summary judgement decision on June 23, 2025. He ruled on three types of uses:

- Granting summary judgment for **Anthropic's copying of works to train LLMs**, ruling that such copying is fair use.
- Granting summary judgment for **Anthropic's conversion of lawfully-purchased print copies to digital** for use in a central library, ruling that such conversion is fair use.
- Denying summary judgment for Anthropic's **copying of "pirated" works for use in the library**, ruling that such copying was not fair use, and allowing that question to proceed to trial.

Training LLMs is fair use

The court weighed the fair use factors in favor of Anthropic on the question of using copyrighted works to train LLMs.

- **Factor 1 – the purpose and character of the use – for Anthropic.** The court found the "purpose and character of using copyrighted works to train LLMs" was "transformative – spectacularly so." Judge Alsup likened what Anthropic was doing to the human act of reading existing texts and writing new texts. "Like any reader aspiring to be a writer, Anthropic's LLMs trained upon works not to race ahead and replicate or supplant them – but to turn a hard corner and create something different." The factor weighed for fair use.
- **Factor 2 – nature of the copyrighted works – against Anthropic.** Anthropic accepted that all the books at issue contained expressive elements. The court accepted that the authors' books were chosen for their expressive qualities. The factor weighed against fair use.
- **Factor 3 – amount and substantiality of the portion used – for Anthropic.** The copies used for training the LLMs were "reasonably necessary to the transformative use" and there was no allegation that Anthropic used the works to generate a public-facing output that used an amount of the original works. The factor weighed for fair use.
- **Factor 4 – effect on the market for or value of the copyrighted works – for Anthropic.** The court found "training LLMs did not result in any exact copies nor even infringing knockoffs of their works being provided to the public" and, thus, training Claude "did not and will not displace demand for copies of Authors' works." The factor weighed for fair use.

Using "pirated" copies to build a central library is not fair use

With respect to the "central library" Anthropic created, Judge Alsup considered how the works were sourced.

The court ruled it was fair use for Anthropic to purchase print copies and then scan them from print to digital for use in the library. This use was transformative, as "every purchased print copy was copied in order to save storage space and to enable searchability as a digital copy. The print original was destroyed. One replaced the other. And, there is no evidence that the new, digital copy was shown, shared, or sold outside the company." The third and fourth factors also favored Anthropic.

Judge Alsup however ruled that it was not fair use for Anthropic to acquire millions of "pirated" copies for use in its central library. The court noted that "not every book Anthropic pirated was used to train LLMs. And, every pirated library copy was retained even if it was determined it would not be so used." This was key: "Pirating copies to build a research library without paying for it, and to retain copies should they prove useful for one thing or another, was its own use – and not a transformative one" The remaining fair use factors weighed against fair use. The court ruled that it will have a trial on the pirated copies used to create Anthropic's central library and any resulting damages.

Takeaways

- The decision is a win for AI companies seeking to establish that copying of works to train AI models is fair use. It establishes the general proposition that AI companies can use copyrighted works to train LLMs.

- But the decision is not a total victory for AI companies. Judge Alsup did not rule on whether using “pirated” works to train LLMs is fair use. In *dicta*, he suggested the opposite: “This order doubts that any accused infringer could ever meet its burden of explaining why downloading source copies from pirate sites that it could have purchased or otherwise accessed lawfully was itself reasonably necessary to any subsequent fair use.” Judge Alsup went on to write: “Such piracy of otherwise available copies is inherently, irredeemably infringing even if the pirated copies are immediately used for the transformative use and immediately discarded.” This reasoning is similar to the U.S. Copyright Office’s views in its pre-publication report on Copyright and Artificial Intelligence.² Judge Alsup ultimately declined to rule on the question, as he found the “pirated” copies infringing for another reason – they were copied for Anthropic’s general-purpose central library.
- Judge Alsup’s decision leaves unanswered the question of whether “outputs” of generative AI products are fair use. That was not before the court, and it awaits a decision elsewhere.

Fair use summary

This summarizes Judge Alsup’s fair use determination:

	Copies used to train LLMs	Purchased library copies converted from print to digital	“Pirated” library copies
Factor 1: purpose and character of the use	Favors Anthropic	Favors Anthropic	Favors authors
Factor 2: nature of copyrighted works	Favors authors		
Factor 3: amount and substantiality of the portions used	Favors Anthropic	Favors Anthropic	Favors authors
Factor 4: effect on market for and value of the works	Favors Anthropic	Neutral	Favors authors
Weighing of factors	Fair use	Fair use	Not fair use

^[1] *Bartz v. Anthropic PBC*, No. 24-cv-05417-WHA (N.D. Cal. June 23, 2025).

^[2] U.S. Copyright Office, *Copyright and Artificial Intelligence – Part 3: Generative AI Training (Pre-Publication Version)* (May 2025) at 52 (“In the Office’s view, the knowing use of a dataset that consists of pirated or illegally accessed works should weigh against fair use without being determinative”).

This informational piece, which may be considered advertising under the ethical rules of certain jurisdictions, is provided on the understanding that it does not constitute the rendering of legal advice or other professional advice by Goodwin or its lawyers. Prior results do not guarantee similar outcomes.

CONTACTS

Robert D. Carroll

Partner

rcarroll@goodwinlaw.com

Boston | +1 617 570 1753

Stefan Mentzer

Partner

smentzer@goodwinlaw.com

New York | +1 212 459 7088

Timothy Keegan

Associate

tkeegan@goodwinlaw.com

New York | +1 646 296 1411

Eunbi Choi

Associate

eunbichoi@goodwinlaw.com

New York | +1 917 228 7833



GOODWIN