Bartz v. Anthropic: Fair Use and Training A.I. on Books

By Thomas A. Guerriero, PhD

Case Commentary | Independent Analysis | Published Wednesday 18 November 2026

Independent commentary by a non-attorney for educational purposes; not legal advice and not filed with any court.

Case: *Bartz v. Anthropic PBC*, No. 3:24-cv-05417-WHA, Order on Fair Use, slip op. (N.D. Cal. June 23, 2025) (Alsup, J.).

Background

Bartz v. Anthropic applied the fair use doctrine to the training of large language models on books. Three authors, Andrea Bartz, Charles Graeber and Kirk Wallace Johnson, together with two companies they set up to market their works, sued Anthropic PBC, the company behind the Claude AI service, for copyright infringement. On an early motion for summary judgment limited to fair use, Judge William Alsup of the Northern District of California split the case in two. Training was fair use. Building a permanent library from pirated books was not. Order on Fair Use, slip op. at 1–2, 8, 31.

The order describes how Anthropic assembled its training material. In January or February 2021 a co-founder downloaded Books3, an online library of 196,640 books he knew had been assembled from unauthorised copies. In June 2021 he downloaded at least five million copies of books from Library Genesis, and in July 2022 Anthropic downloaded at least two million copies from the Pirate Library Mirror, both of which, the order states, Anthropic knew had been pirated. *Id.* at 2–3.

Anthropic later changed course. It spent many millions of dollars buying millions of print books, often used. Its service providers stripped the bindings, cut the pages, scanned them into digital form and discarded the paper originals. *Id.* at 4. The pirated and scanned copies were kept in a central library that Anthropic intended to retain "forever." From that library, Anthropic selected sets and subsets of books to train the models underlying Claude. *Id.* at 1–2.

Two features of the record shaped what followed. First, the authors did not allege that Claude produced infringing copies of their works for users; the order records that Anthropic used filtering software on inputs and outputs and that the authors challenged "only the inputs, not the outputs" of the models. *Id.* at 7, 11–12. Second, for the purposes of the motion, the order took as given the authors' contention that the trained models had memorised the works almost verbatim. *Id.* at 7.

Holding

The court granted summary judgment for Anthropic that the training use was fair use, and that converting purchased print books into digital copies was fair use for a different reason. It denied summary judgment that the pirated library copies must be treated as training copies, and set a trial on the pirated copies used to create the central library and on the resulting damages. *Id.* at 31.

The order also left questions open. It did not grant summary judgment as to any copies made from the central library but not used for training, finding the record too poorly developed. *Id.* at 31–32.

Reasoning

Training as transformative use

The court treated the purpose and character of the use as decisive for training. It concluded that "the purpose and character of using copyrighted works to train LLMs to generate new text was quintessentially transformative." The order analogised training to a reader who aspires to be a writer, learning from works in order "to turn a hard corner and create something different." *Id.* at 13–14.

The other factors followed. The second factor, the nature of the works, pointed against fair use for all copies alike. *Id.* at 24. On amount, the court accepted that copying entire books was reasonable for training and noted, again, that the authors alleged no traceable connection between Claude's outputs and their works. *Id.* at 25. On market effect, the court rejected the argument that training would produce an explosion of competing works. In a passage that later drew criticism from another judge in the same district, it said the authors' complaint was "no different than it would be if they complained that training schoolchildren to write well would result in an explosion of competing works." *Id.* at 28. Weighing the factors, the court concluded that the training copies were justified and that every factor but the nature of the works favoured that result. It described the technology as "among the most transformative many of us will see in our lifetimes." *Id.* at 30.

Format change for purchased books

The court treated the scanning of purchased print books separately. It held this was fair use because Anthropic replaced print copies it had bought with more convenient digital copies, destroyed the print originals and did not distribute the digital copies outside the company. *Id.* at 9, 15, 30–31. This was a narrower justification than training. It did not depend on the copies being used to build an AI system.

Pirated library copies

The court reached the opposite conclusion for the pirated copies. Anthropic argued that downloading them was justified because the copies were reasonably necessary for training. The court disagreed. It observed that piracy of otherwise available copies could be seen as "inherently, irredeemably infringing" even if the copies were immediately used for a transformative purpose and discarded, but expressly said it need not decide the case on that rule. *Id.* at 19. Instead, it found that Anthropic's use was building a central library of works for any number of further uses, that not every pirated book was used for training, and that pirated copies were retained even when Anthropic had decided not to use them for training. That use was not transformative. *Id.* at 19.

On market effect, the court held that stealing copies that could have been bought harmed the market for the authors' works, noting that Anthropic could have purchased the authors' books and later did. *Id.* at 30. Every factor pointed against fair use for the pirated library copies. *Id.* at 31. The fact that Anthropic later bought copies of books it had earlier downloaded would not absolve it of liability, although it might affect statutory damages. *Id.* at 31–32.

Significance

The order is often summarised as holding that AI training is fair use. That summary is incomplete. The decision distinguished sharply between what was done with copies and how the copies were obtained and kept. Training was protected on this record. Acquiring pirated books to build a general-purpose library was not, and that conclusion exposed Anthropic to a trial on statutory damages.

The decision was also limited by its record. The authors challenged inputs, not outputs, and alleged no infringing outputs reached users. The order said that if outputs were infringing, the authors "would have a different case." *Id.* at 12. It was an order of a single district judge, not an appellate ruling.

Later proceedings, verified from court filings, changed the practical significance of the case. The court certified a class of copyright owners of books in the versions of Library Genesis and the Pirate Library Mirror that Anthropic downloaded. Order on Class Certification at 31 (N.D. Cal. July 17, 2025). It denied Anthropic's motion to stay proceedings pending appellate review, with trial then set for 1 December 2025. Order Denying Motion to Stay at 1 (N.D. Cal. Aug. 11, 2025). The parties then settled. In a memorandum opinion explaining preliminary approval, the court described a per-work award of about $3,000, less costs and fees, an injunction requiring destruction of copies derived from Library Genesis and the Pirate Library Mirror, and a release limited to past conduct concerning AI inputs, not outputs. Preliminary Approval Opinion at 5 (N.D. Cal. Oct. 17, 2025). On 20 July 2026, Judge Araceli Martínez-Olguín granted final approval of a non-reversionary settlement fund of $1.5 billion, entered final judgment and dismissed the action with prejudice. Final Approval Order at 1, 4, 23 (N.D. Cal. July 20, 2026). This commentary has found no appellate ruling on the fair use order, and the action has now been dismissed under the settlement.

Practical Takeaways

The following are general governance observations drawn from the decision, not legal advice about any particular use of data.

- **Separate acquisition from use.** The court analysed how copies were obtained and kept apart from how they were used. A lawful use did not cure an unlawful acquisition.
- **Know the provenance of training data.** Records showing where material came from, and whether it was licensed or purchased, matter.
- **Avoid indefinite retention without purpose.** Keeping copies "forever" for general purposes weighed heavily against fair use.
- **Monitor outputs.** The authors' case did not involve infringing outputs; the court said such outputs would present a different case.
- **Read the later history.** The fair use ruling and the $1.5 billion settlement concern different questions; neither should be cited for more than it decided.

These observations draw on pages 1–4, 7, 12–14, 19, 28–32 of the order and the later orders cited above. They are offered to support careful data governance and informed discussion with qualified counsel.

Open Questions

The order leaves important questions open. It did not decide whether infringing outputs would defeat fair use for training. It did not decide how copies made from the central library for purposes other than training should be treated. And with the case resolved by settlement, there is no appellate ruling in it on whether training on lawfully acquired books is fair use. In *Kadrey v. Meta Platforms*, decided two days later in the

same court, Judge Vince Chhabria criticised the schoolchildren analogy and placed greater weight on potential market dilution. *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC, slip op. at 3 (N.D. Cal. June 25, 2025).

For organisations developing or using AI tools, the most durable lesson of *Bartz* may be that the source of training material can matter as much as the use made of it.

References

- United States District Court for the Northern District of California, *Bartz v. Anthropic PBC*, No. 3:24-cv-05417-WHA, Order on Fair Use (filed 23 June 2025), official U.S. Courts opinion published on GovInfo: https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_24-cv-05417/pdf/USCOURTS-cand-3_24-cv-05417-0.pdf
- Same court, Order on Class Certification (17 July 2025): https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_24-cv-05417/pdf/USCOURTS-cand-3_24-cv-05417-1.pdf
- Same court, Order Denying Motion to Stay (11 August 2025): https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_24-cv-05417/pdf/USCOURTS-cand-3_24-cv-05417-4.pdf
- Same court, Memorandum Opinion on Preliminary Approval of Class Action Settlement (17 October 2025): https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_24-cv-05417/pdf/USCOURTS-cand-3_24-cv-05417-5.pdf
- Same court, Order Granting Final Approval of Class Action Settlement; Judgment, Dkt. No. 680 (20 July 2026), court filing as reproduced by Justia: https://cases.justia.com/federal/district-courts/california/candce/4:2024cv05417/434709/680/0.pdf
- Same court, *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC (25 June 2025): https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_23-cv-03417/pdf/USCOURTS-cand-3_23-cv-03417-37.pdf