

Bartz v. Anthropic PBC
787 F. Supp. 3d 1007 (N.D. Cal. 2025)

Year	2025
Court	United States District Court for the Northern District of California
Key Facts	Defendant, Anthropic PBC, is a generative AI software company that develops and deploys large language models (LLMs) in its products, including Claude, which provide textual responses to user prompts. To train its various LLMs, Anthropic allegedly downloaded millions of copyrighted books from pirate websites. Anthropic also purchased books, mostly used copies, in bulk, which it then deconstructed and scanned. Anthropic allegedly maintained copies of both pirated and purchased books in its “research library” as a permanent, general resource. Anthropic’s engineers also allegedly copied books during training when creating “data mixes,” and when they were cleaned and tokenized. Some compressed copies allegedly may have been retained in the LLMs. Plaintiffs are authors—representative of potential classes of authors—who allege that their books were copied from the pirated and/or purchased sources. Plaintiffs sued Anthropic, alleging copyright infringement, but did not allege that any infringing content was provided to Claude users in response to prompts. Anthropic moved for summary judgment, asserting fair use.
Issue	Whether using copies of pirated and purchased books to create an internal library and to train publicly deployed generative AI models are fair uses.
Holding	In conducting the fair use analysis, the court analyzed the four factors separately as to (a) training and (b) building a research library. The court also distinguished between lawfully purchased and pirated copies in its analysis. Considering the first factor, the purpose and character of the use, the court found that Anthropic’s use of the copies to train LLMs was “quintessentially transformative.” The court noted that Plaintiffs did not allege that any of Claude’s outputs infringed their works and that Anthropic used software to ensure that no infringing material reached users. The court analogized LLM training to a reader aspiring to be a writer who consumes and memorizes many books as inspiration to create a new work. As for the use of works to create a library, the court found that converting purchased books from print to digital was transformative because it saved storage space, enabled searchability, and did not result in the distribution of new copies. However, the court found that downloading and retaining pirated copies in a central library was not transformative, even where Anthropic subsequently used those works in model training or paid for authorized copies. The second factor, the nature of the copyrighted work, weighed against fair use for all copies, whether purchased or pirated, because “[Plaintiffs’] works were chosen for their expressive qualities.” The third factor, the amount and substantiality of the portion used, favored fair use for LLM training. The court reasoned that, while Anthropic copied entire books and did not need any particular work, using large quantities of high-quality texts was reasonably necessary for the transformative purpose of LLM training and no copies were publicly revealed. As for using the works to build a library, the third factor pointed in favor of fair use for the purchased copies because digitizing entire works was reasonable to facilitate storage and searchability, and copying each work in full was necessary to achieve the purpose. But this factor pointed against fair use for the pirated materials as Anthropic “lacked any entitlement to hold copies of the books” indefinitely “on the chance that some might prove useful.”

	Finally, the court concluded that the fourth factor, the effect upon the potential market for or value of the original works, favored fair use for training LLMs because that use did not result in any outputs that would harm the market for the originals. The court concluded “[t]he copies used to train specific LLMs did not and will not displace demand for copies of [Plaintiffs’] works.” The court declined to recognize any potential market harm attributable to “an explosion of competing works” made possible by use of LLMs or for displaced opportunities to license the works for LLM training. The court found the fourth factor was neutral for the purchased library copies because the format change from print to digital, absent any further dissemination of the works, “did not itself usurp [Plaintiffs’] rightful entitlements.” The factor weighed against fair use for the pirated library copies because permitting Anthropic’s conduct, “steal[ing] a work you could otherwise buy” while “loosely intend[ing] to make future copies for a purportedly transformative purpose,” would “destroy” the market for such works. In conclusion, weighing the factors together, the court held that using copies to train LLMs was “justified as fair use.” The court determined that Anthropic’s use of purchased physical copies to build a digital library was likewise fair use overall because the digital replacement copies were not redistributed. On the other hand, the court held Anthropic’s use of pirated copies to build its digital library was not fair because all four factors weighed against fair use.
Tags	Internet/Digitization; Computer Program
Outcome	Mixed Result

Source: U.S. Copyright Office Fair Use Index. For more information, see <https://www.copyright.gov/fair-use>.