Kadrey v. Meta Platforms, Inc.
788 F. Supp. 3d 1026 (N.D. Cal. 2025)

Year	2025
Court	United States District Court for the Northern District of California
Key Facts	Defendant Meta Platforms, Inc. ("Meta") downloaded datasets consisting of books from "shadow libraries" (*i.e.*, online repositories that provide free access to media regardless of whether it is copyrighted) through a torrenting network and used them to train various versions of its generative AI large language model (LLM), Llama. Meta attempted to license books from publishers but ultimately abandoned this method of book acquisition "because Meta realized that licensing would be more difficult than anticipated" for several reasons. Meta also post-trained its LLMs to prevent them from memorizing and outputting a certain amount of text. Plaintiffs, thirteen famous authors whose works can be found in the datasets, sued Meta for copyright infringement (among other claims, including secondary liability and DMCA violations) based on Meta's reproduction and alleged distribution of Plaintiffs' books. Following partial dismissal of other claims, both Plaintiffs and Meta moved for partial summary judgment as to whether Meta's reproduction constituted infringement, with Meta asserting fair use.
Issue	Whether using books obtained from internet shadow libraries for generative AI training purposes constitutes fair use.
Holding	Considering the first factor, the purpose and character of the use, the court found that Meta's use was "highly transformative." The court reasoned that the purpose of training LLMs, as "innovative tools that can be used to generate diverse text and perform a wide range of functions," is distinct from the purpose of the books, which is "to be read for entertainment and education." The court further observed that LLMs "consum[e]" books to learn statistical patterns about how words are used and that is different from how humans read and learn. It also noted that Llama did not regurgitate any works, only certain "styles" of text, commenting that "style is not copyrightable." As for commerciality, the court found that while Llama is available under a free license, Meta developed it for commercial reasons and stands to make significant profits over the next decade, though this was outweighed by the use's transformativeness. The court noted the use of shadow libraries was relevant to bad faith and could affect the character of the use, but didn't "move the needle" in this case. On balance, the court found the first factor favored fair use as to LLM training (and downloading works for that purpose). The second factor, the nature of the copyrighted work, weighed against fair use because Plaintiffs' books are "highly expressive" and Meta's use depended on that expression; but this factor was not significant in the overall fair use analysis and applied "with less force" because the copied works were already published. Next, the court noted that the third factor, the amount and substantiality of the portion used, was not "especially relevant" here because Meta's LLMs were unlikely to output a "meaningful amount" of Plaintiffs' books. The court went on to find that the factor favored fair use even though Meta copied the entirety of Plaintiffs' works because "LLMs work better if they are trained on more high-quality material," so that amount of copying was reasonably necessary to achieve the transformative purpose. Finally, for the fourth factor, the effect upon the potential market for or value of the original works, the court considered three theories of market harm as well as potential public benefits before concluding that the factor favored fair use. The court first concluded that Meta's use did not result in direct market substitutes because Llama did not output a meaningful amount of Plaintiffs' works in response to user prompts. Second, the

	court rejected Plaintiffs' argument that "unauthorized use of their books for LLM training harms the market for licensing their books for that purpose," concluding that, even if such market existed, Plaintiffs are not "legally entitled to monopolize" it. Third, the court addressed "market dilution," commenting that the harm that an LLM poses in its ability to generate numerous competing, but non-infringing, outputs would differ depending on the type of work. The court concluded that such "indirect substitution" was potentially cognizable and "highly relevant" to this type of copyright dispute but was insufficiently pleaded in this case. In addition, the court noted the public benefits of Meta's use, which "will likely help . . . create new expression," favor fair use. The court, however, rejected Meta's suggestion that requiring model developers to pay to use copyrighted text as training data would disserve the "public interest." In conclusion, the court held that absent evidence of market dilution, "on balance" the factors favored fair use and Meta was entitled to summary judgment.
Tags	Internet/Digitization; Computer Program
Outcome	Fair use found

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