

How Federal Courts Are Applying the Fair Use Doctrine to AI

July 2, 2025

Federal courts have recently clarified the application of the fair use doctrine to the use of copyrighted works in training large language models (LLMs) for artificial intelligence (AI). These decisions set important boundaries for AI developers, indicating that while the act of training may be protected as fair use, the method of acquiring training data—especially through piracy—remains subject to liability.

I. *Bartz v. Anthropic PBC*

In *Bartz v. Anthropic PBC*, authors sued Anthropic for copying their works, both pirated from online sources and purchased in print and then digitized, to build a central digital library and to train its AI software service, Claude. [1] Anthropic downloaded millions of pirated books from sites like Books3 and LibGen, and also purchased and physically destroyed millions of print books to scan them and create a searchable digital library. [2] The company used subsets of these books to train various LLMs, retaining all copies in its central library for potential future uses. [3] The plaintiffs alleged copyright infringement for both the creation of the central library and the use of their works in LLM training but did not allege that the LLM outputs infringed upon their works. [4] In response, Anthropic argued that pirating initial copies of these works was justified because these copies were at least reasonably necessary for training LLMs. [5]

To determine whether the alleged infringement was defensible, the court undertook a factor-by-factor analysis of fair use under Section 107 of the Copyright Act.

1. Purpose and Character of the Use

First, in assessing the purpose and character of the use, the court held that using copyrighted works to train LLMs was a “spectacularly” transformative fair use, as the training process was “orthogonal” to the original purpose of the works and did not result in infringing outputs to the public. [6] However, the court found that simply building and retaining a permanent central library of pirated copies was not considered transformative and weighed against fair use. [7]

2. Nature of the Copyrighted Work

Next, the court recognized that the works at issue, published fiction and nonfiction books, were highly expressive and thus “closer to the core of intended copyright protection...” [8] Both the central library and the training sets

included works chosen for their creative qualities, and the court accepted that these expressive elements were valued by Anthropic in building its datasets. This factor therefore weighed against fair use for all copies, as the works were not merely factual or utilitarian, but the court noted that this factor primarily serves to inform the analysis of the other factors.[9]

3. Amount and Substantiality of the Portion Used

Third, the court found that Anthropic copied entire works for both its central library and for training LLMs, but distinguished between the uses. For training, the court held that copying the full works was “especially reasonable” given the transformative purpose and the lack of evidence that the LLMs’ outputs substituted for the original works.[10] The court noted that the volume of text required to train an LLM is monumental, and that using many works was “reasonably necessary” for the transformative use.[11] In contrast, for the pirated library copies, the court found that copying millions of books for the purpose of building a general-purpose library was not justified, as this was not a transformative use and almost any unauthorized copying for this purpose would have been too much.[12]

4. Effect of the Use on the Potential Market

For the training copies, the court held that there was no evidence that the use displaced demand for the plaintiffs’ works as contemplated by the Copyright Act, as the LLMs did not output infringing copies or substantial knock-offs to the public.[13] The court rejected the argument that the potential for LLMs to generate competing works, such as alternative summaries or examples of writing, constituted the kind of market harm the Copyright Act is concerned with, likening it to the effect of teaching schoolchildren to write well.[14] The court also found that the emergence of a licensing market for AI training was not a market the Copyright Act entitles authors to control, and thus loss of such a market did not weigh against fair use.[15] For the pirated library copies, however, the court found that the acquisition of unauthorized copies plainly displaced demand for the plaintiffs’ books “copy for copy,” and that condoning such conduct would destroy the publishing market.[16]

5. Overall Conclusion and Holding

In sum, the court held that the use of copyrighted works to train LLMs was a transformative fair use, with the first and third factors strongly favoring Anthropic, the second factor weighing against, and the fourth factor favoring fair use for training but not for the creation of a permanent library from pirated copies. The court granted summary judgment for Anthropic on the fair use defense for training and format conversion, but denied it as to the pirated library copies, which would proceed to trial for damages.[17]

II. *Kadrey v. Meta Platforms, Inc.*

In *Kadrey v. Meta Platforms, Inc.*, thirteen authors sued Meta for downloading their copyrighted books from online “shadow libraries” and using them to train its Llama large language models without permission.[18] Meta initially attempted to license books from publishers but, after facing logistical challenges, resorted to downloading books from unauthorized sources such as LibGen and Anna’s Archive.[19] The plaintiffs alleged that Meta’s actions constituted copyright infringement and harmed the market for their works, seeking damages and

injunctive relief.[20] Meta did not dispute the copying but argued that its use was transformative and protected by the fair use doctrine.[21]

The court analyzed the fair use factors:

1. Purpose and Character of the Use

The court found Meta’s use of the plaintiffs’ books to train its Llama AI models was “highly transformative,” as the purpose was to develop a tool capable of generating diverse text and performing a wide range of functions, distinct from the original purpose of the books.[22] Although Meta’s use was commercial, the transformative nature of the use outweighed the commercial aspect, and the court noted that commercialism is less important when the use is highly transformative.[23] The court also rejected arguments that bad faith or the manner of acquisition (downloading from shadow libraries) automatically precluded fair use, finding these issues were not determinative in light of the transformative purpose.[24]

2. Nature of the Copyrighted Work

The court determined this factor favored the plaintiffs, as their works—mostly novels, memoirs, and plays—are highly expressive and at the core of copyright protection.[25] However, the court noted that this factor is generally less significant in the overall fair use analysis, especially since the works had already been published.[26] Meta’s argument that it only used the “functional elements” of the works was rejected, as the AI models benefit from the creative expression in the books.[27]

3. Amount and Substantiality of the Portion Used

Although Meta copied the plaintiffs’ books in their entirety, the court found this was reasonable and necessary for the transformative purpose of training an effective AI model.[28] The court emphasized that the relevant consideration is not the amount copied, but the amount made available to the public, and here, the AI models did not output meaningful portions of the plaintiffs’ works.[29] This factor, therefore, favored Meta.

4. Effect of the Use on the Potential Market

The court held this is the most important factor and found for Meta, as the plaintiffs failed to present evidence that Meta’s use caused or was likely to cause market harm to their works.[30] The court rejected the plaintiffs’ arguments regarding loss of licensing markets for AI training and the risk of the AI regurgitating their works, finding no evidence of significant market substitution or dilution.[31] While the court acknowledged that market dilution from AI-generated competing works could be a viable theory in other cases, the plaintiffs here did not develop the record or present empirical evidence to support such harm.[32]

5. Overall Conclusion and Holding

The court concluded that, although Meta’s use was highly transformative and the plaintiffs’ works were at the core of copyright protection, the lack of evidence of market harm was dispositive. As a result, Meta prevailed on its fair use defense to the claim that copying these plaintiffs’ books for use as LLM training data was

infringement.[33]

III. Conclusion

These holdings will have a profound impact on the development and deployment of AI by providing clearer guidance on the boundaries of fair use in AI training.

Cullen and Dykman’s Intellectual Property team continues to monitor important developments in trademark and copyright law. Should you have any questions about this legal alert, please feel free to contact Karen Levin (klevin@cullenllp.com) at (516) 296-9110, Ariel Ronneburger (aronneburger@cullenllp.com) at (516) 296-9182, or Jordan Milite (jmilite@cullenllp.com) at (516) 296-9128.

This advisory provides a brief overview of the most significant changes in the law and does not constitute legal advice. Nothing herein creates an attorney-client relationship between the sender and recipient.

Footnotes

[1] *Bartz et al. v. Anthropic PBC*, No. 3:24-cv-05417-WHA (N.D. Ca. June 23, 2025).

[2] *Id.* at 2-4, 7.

[3] *Id.* at 5.

[4] *Id.* at 7, 11.

[5] *Id.* at 8.

[6] *Id.* at 11-13 (citing *Andy Warhol Found.for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 538-540 (2023)).

[7] *Bartz*, No. 3:24-cv-05417-WHA at 19.

[8] *Id.* at 24 (citing *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 586 (1994)).

[9] *Bartz*, No. 3:24-cv-05417-WHA at 24.

[10] *Id.* at 25-26.

[11] *Id.* at 26.

[12] *Id.* at 27.

[13] *Id.* at 28.

[14] *Id.* at 28 (citing *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1523–24 (9th Cir. 1992)).

[15] *Id.*

[16] *Id.* at 29-30.

[17] *Id.* at 30-32.

[18] *Richard Kadrey, et al. v. META Platforms, Inc.*, No. 23-cv-03417-VC, p. 4, 11 (N.D. Ca. June 25, 2025).

[19] *Id.* at 10-13, 38.

[20] *See id.* at 33.

[21] *See id.* at 16.

[22] *Id.* at 15-17 (citing *Warhol*, 598 U.S. at 528 and *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021)).

[23] *Id.* at 18-19.

[24] *Kadrey*, No. 23-cv-03417-VC at 19-20.

[25] *Id.* at 23 (citing *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 98 (2d Cir. 2014)).

[26] *Kadrey*, No. 23-cv-03417-VC at 24 (citing *Google Books*, 804 F.3d at 220).

[27] *Id.* at 23.

[28] *Id.* at 25.

[29] *Id.* (quoting *Fox News Network, LLC v. TVEyes*, 883 F.3d 169, 179 (2d Cir. 2018)).

[30] *Kadrey*, No. 23-cv-03417-VC at 34-36.

[31] *Id.* at 27-28; 38.

[32] *Id.*

[33] *Id.* at 40.

Practices

- Intellectual Property

Attorneys

- Ariel E. Ronneburger
- Karen I. Levin
- Jordan Milite