

Publications

Fair Use or Infringement? Recent Court Rulings on AI Trained on Copyrighted Works

Related Attorneys

William H. Oldach III

Athena J. Williams

Related Services

Copyrights

Intellectual Property

CLIENT ALERT | 8.28.2025

Artificial intelligence (AI) is now part of daily life, powering customer service chatbots, virtual assistants like Siri and Alexa, automated email responses, and personalized shopping recommendations. But as these systems get smarter, they need ever-larger amounts of data to learn, often drawing on copyrighted books and creative works. This has led to new legal battles over whether AI companies are crossing the line into copyright infringement, or whether their use of these materials to train large language models qualifies as “fair use.”

Meta’s Fair Use Win: The Importance of Market Impact

One of the most closely watched cases in this area is *Kadrey v. Meta Platforms, Inc.*, involving Meta (Facebook’s parent company) and its AI model “Llama.” The authors who brought the lawsuit argued that Meta used their books to train its AI without permission and that this would harm their ability to license their works in the future. The court approached the issue by carefully considering the four “fair use” factors: the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality used, and the effect on the market for the original.

Crucially, the court found that Meta’s use was “transformative,” meaning that the books were not simply republished or copied but instead were used to teach an AI how to generate human-like language. The decision ultimately focused on whether Meta’s AI-generated texts would flood the market and hurt sales of the copyrighted books. The authors argued that the AI could reproduce snippets from their works and harm their licensing opportunities, but the court found that there was not enough evidence that Meta’s use actually damaged book sales or the authors’ market. As a result, Meta prevailed, though the court noted that future cases with stronger evidence of a negative market impact could reach a different outcome.

Anthropic: Class Action, Copyright Risks, and a Settlement

A very different story unfolded in *Bartz v. Anthropic PBC*. Anthropic, the company behind the AI chatbot “Claude” was sued by a group of authors who claimed that Anthropic had illegally downloaded millions of copyrighted books from pirate sites like Library Genesis (LibGen) and Pirate Library Mirror to train its AI. The scale of the alleged infringement was massive, and the authors sought to represent not just themselves, but a class comprising every copyright owner whose work was scraped from these sources.

In a June 2025 ruling regarding fair use, the court found that using books to train an AI could be “transformative”—meaning it was a new use that did not simply replace the original books. As long as the AI did not spit out large chunks of the original books to users, this kind of use might be considered fair under copyright law. However, the court drew a line: simply downloading and storing pirated books in a database, without transforming them or using them for a new purpose, was not fair use. The act of piracy itself was not excused just because the books might eventually be used for something transformative.

On July 17, 2025, the court allowed the case to move forward as a class action, meaning all affected authors and copyright holders of pirated books could pursue their claims together. The court found this approach fair and efficient, especially given the large scale of the alleged infringement. However, the class was limited to works downloaded from LibGen and Pirate Library Mirror, and the court established a process for determining which works were obtained legally and which were pirated, acknowledging the complexity involved in identifying class members and matching them to specific works.

Now, the *Anthropic* case is nearing its conclusion. According to recent court filings, Anthropic and the authors have reached a proposed settlement. The parties have asked the court to pause all proceedings while they finalize the details, and they expect to file a motion for approval of the settlement by early September. While the full terms of the settlement have not yet been made public, attorneys for the authors have described it as “historic” and beneficial for all class members. This resolution could set an important precedent for how future copyright disputes involving AI are handled, signaling that large-scale, unauthorized use of copyrighted works to train AI can lead not only to significant legal exposure, but also to meaningful remedies for creators.

What Does This Mean for You?

If you are an author or creator, these cases show that your work might be used to train AI models without your knowledge, but they also demonstrate that legal remedies are available, especially if your work was pirated and used by a tech company. For businesses and startups, the lesson is clear: use only properly licensed training data, as relying on pirated or unlicensed material can result in expensive lawsuits and reputational harm. Consumers should expect more legal battles over who owns the content they see online and how it was created, since some AI-generated text, images, or music may be based on copyrighted works, raising important questions about originality and ownership.

As the law catches up to the rapid growth of AI, courts are beginning to set boundaries for how creative works can be used to power new technology. The outcomes of these cases, and settlements like Anthropic’s, will help define what’s fair, what’s legal, and what’s possible in the age of artificial intelligence.