

CLIENT ALERTS

AI and Copyright: How a Recent AI-Related Decision May Impact Your AI Strategy and Data Practices

Client Alert

6.26.2025

As artificial intelligence (AI) tools become more integrated into day-to-day operations—from content generation to customer service—businesses are increasingly relying on technologies that may be trained on massive datasets pulled from the internet.

A recent court order in *Bartz et al. v. Anthropic PBC* underscores a growing legal risk: if those underlying datasets include copyrighted materials without permission, companies using or deploying these AI systems could face exposure. Whether you are building AI, licensing it, or simply using it in your business, this case is a wake-up call to reassess your contracts, compliance practices, and risk management strategies.

Brief Background:

In the *Anthropic* case, a group of authors filed a lawsuit alleging that Anthropic, a company that develops large language models (LLMs), built AI models that were trained on copyrighted works without permission. The plaintiffs alleged that some of the works (books) were purchased while others were pirated to build a library and fed into large language models without any license or compensation.

Anthropic argued that its use of text to train AI systems was protected by the doctrine of fair use, a legal standard that allows limited use of copyrighted works without permission for purposes like teaching, research, or transformative uses.

On June 23, 2025, Judge William Alsup of the US District Court for the Northern District of California issued an order, partially siding with Anthropic.

Related People

Erin Malone
Associate

Claudia Rast
Shareholder

Maya Smith
Associate

Andrew S. AbdulNour
Associate

Related Services

Artificial Intelligence

Copyright

Intellectual Property

CLIENT ALERTS

The court walked through the four fair use factors in its analysis:

- (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- (4) the effect of the use upon the potential market for or value of the copyrighted work.

Key Issues:

The court examined three key issues:

1. Digitization of Books: The court provided that the digitization of books *purchased* in print form by Anthropic was also fair use because essentially all Anthropic did was replace print copies it had purchased for its central library with more convenient space-saving and searchable digital copies for its central library without adding new copies, creating new works, or redistributing existing copies.

“Here, every purchased print copy was copied in order to save storage space and to enable searchability as a digital copy. The print original was destroyed. One replaced the other. And, there is no evidence that the new, digital copy was shown, shared, or sold outside the company.”

2. Training Copies: The court also analyzed the use of the books for training AI models. The authors did not allege that the training process resulted in exact copies or infringing outputs of the authors’ works being made available to the public. This was an important point for the court. The court then instead solely considered the inputs of the LLMs and whether the use of copyrighted material for this purpose was transformative and reasonable under fair use principles. The court ultimately called the use to train the LLM “exceedingly transformative.”

“[I]f someone were to read all the modern-day classics because of their expression, memorize them, and then emulate a blend of their best writing, would that violate the Copyright Act? Of course not. Copyright does not extend to ‘methods of operation, concepts, or principles illustrated or embodied in a work.’ . . . In short, the purpose and character of using copyrighted works to train LLMs to generate new text was quintessentially transformative. Like any reader aspiring to be a writer, Anthropic’s LLMs trained upon works not to race ahead and replicate or supplant them—but to turn a hard corner and create something different.”

“But Authors’ complaint is no different than it would be if they complained that training schoolchildren to write well would result in an explosion of competing works. This is not the kind of competitive or creative displacement that concerns the Copyright Act. The Act seeks to advance original works of authorship, not to protect authors against competition.”

CLIENT ALERTS

3. Pirated Library Copies: Anthropic had downloaded millions of pirated books to create a central library. The court found that this act of copying was not justified under the fair use doctrine, as it substituted for purchasing legitimate copies and undermined the market for the authors' works.

"For the pirated library copies, however, Anthropic lacked any entitlement to hold copies of the books at all. Its purpose, it says, was to train LLMs. But its objective conduct was to seek 'all the books in the world' and then retain them even after deciding it would not make further copies from them for training – indicating there were other further uses. Against the purpose of acquiring all the books one could on the chance some might prove useful for training LLMs and maybe other stuff too, almost any unauthorized copying would have been too much. Anthropic copied millions of books in toto, Authors' works among them."

The order granted judgment for Anthropic that the print-to-digital format change was fair use and the training use was fair use. However, the court reserved the issues of the pirated copies for trial.

Key Takeaways and Implications:

- Companies developing AI systems must carefully evaluate whether their use of copyrighted works qualifies as fair use, considering factors such as the purpose of the use, the nature of the works, the amount used, and the potential market impact.
- The case underscores the importance of obtaining proper licenses or permissions when using copyrighted material, especially for AI training purposes.
- Maintaining transparency and proper documentation of data sources is critical to avoid legal disputes.
- Businesses should never use pirated materials. The court made it abundantly clear that building a large "just in case" library of pirated works is illegal.
- Implement strict source controls. Businesses are wise to document legitimate acquisition and purge any questionable data.

For businesses involved in AI development, this case serves as a reminder of the legal and ethical considerations surrounding the use of third-party content. If you are using or planning to use copyrighted material in your operations, it may be worth reviewing your practices to ensure compliance with copyright laws. The ruling also is an indication that courts addressing these issues, such as the Northern District of California in this case, may permit training AI models on properly licensed content in certain circumstances. The specifics of those circumstances will vary. Contact the authors of this Alert or your Butzel attorney for more information.

Erin Malone

313.225.7063

malone@butzel.com

CLIENT ALERTS

Claudia Rast

734.213.3431

rast@butzel.com

Maya Smith

313.983.7495

smithmaya@butzel.com

Andrew S. AbdulNour

734.213.3251

abdulnour@butzel.com