CLIENT ALERTS & INSIGHTS

Judge Certifies Authors' Copyright Class Action Against Anthropic

JULY 23, 2025

Authors: Steven M. Selna, Carlo Lipson

On July 17, 2025, US District Court Judge William Alsup approved a class certification against Anthropic for copyright infringement. According to Judge Alsup, it will be straightforward for the entire class to prove harm caused by Anthropic's "Napster-style downloading of millions of works" from pirate libraries and that "if not brought as a class action, there will likely be no action at all."

This decision comes several weeks after Judge Alsup ruled that Anthropic's use of legally purchased books to train its large language model (LLM) constitutes Fair Use under the Copyright Act. However, Anthropic's means of acquiring other copyrighted works by downloading source copies of pirated materials is *not* excused by their ultimate purpose of LLM training.

Following that ruling, Anthropic filed an interlocutory appeal on July 14, 2025, seeking the Ninth Circuit's guidance on two controlling questions:

1. Whether fair use is analyzed based on the defendant's ultimate purpose in using a copyrighted work or instead parsed into separately analyzed constituent steps; and
2. Whether a defendant's acquisition of a copyrighted work from a third party that distributed it without permission strongly weighs against the availability of the fair use defense even if the use is otherwise fair.

Factual Background

Pirate libraries are open repositories of books, digital media, and academic papers that are usually copyrighted or paywalled.

In 2021, Anthropic downloaded over 196,000 unauthorized copies of copyrighted books from Books3, a pirate e-library that "conveniently packaged for mass download pairs of each book's extracted text and filename, enabling the books to be readily rebuilt into separate files or reviewed." Seeking more training material, Anthropic turned to LibGen, another pirate library. Anthropic's co-founder used the BitTorrent protocol, which enables decentralized electronic peer-to-peer file sharing across multiple sources, to mass download five million eBook files.

Third parties then copied the LibGen library to create Z-Library, which was itself copied to create PiLiMi (Pirate Library Mirror), before being shut down by the government in November 2022. In July 2022, Anthropic compared the LibGen and PiLiMi libraries, torrenting two million copies of works not already in its training library.

Anthropic admits it downloaded copyrighted works from LibGen and PiLiMi, but claims doing so is permitted under the Fair Use doctrine. In objecting to the class certification, Anthropic also argued that errors in torrenting led to partial book downloads, making their infringement *de minimis*. In response, Judge Alsup observed that the company was unable to provide a single example of a partial book in the LibGen downloads, and in any case, "stealing a page of copyrighted work is still a violation."

LLM Training Techniques Will Help Identify Class Members

The metadata and hashing techniques Anthropic used to identify copied works for LLM training will assist in identifying eligible class members.

Anthropic used these methods to mitigate a decline in LLM performance. The company realized its LLM, Claude, was trained on multiple copies of certain books, causing the model to memorize the input text. This undermined Claude's performance and presented a risk that the model would produce outputs containing portions of the copyrighted training material.

However, when Claude stopped being trained on pirated books, developers noticed another "performance hit." Anthropic sought training materials of similar quality to the pirated books, opting to download catalogs of metadata from the books it torrented from LibGen and PiLiMi. This metadata contained Industry Standard Book Numbers (ISBN) and Amazon Standard Identification Numbers (ASIN), which can be used to quickly identify books and unpack "intelligence about editions and languages." Anthropic's own records show that commercial metadata (ISBN/ASIN) is generally unique to books at the edition-level, alleviating concerns that different works and editions often share titles and authors.

The metadata catalogs also contained hash values, or short alphanumeric signatures, which algorithms can use to decipher differences in source file content, formatting, or file type. Plaintiffs are using this metadata to identify beneficial owners of copyright registration certificates who are eligible to join the class action.

Class Certification Granted for eBooks but Denied for Scanned Books

Judge Alsup limited the class to actual and beneficial copyright owners of books with an ISBN or ASIN that were downloaded from two pirate libraries: LibGen or PiLiMi. He declined to include the third pirate library, Books3, reasoning that identifying titles and authors could prove too difficult given that these downloads had more incomplete files and comparatively less metadata.

Judge Alsup declined to certify a class of authors of physical books that Anthropic bought and scanned beginning in 2024.

Litigation Timeline and Near-Term Developments

By August 1, 2025, Anthropic must submit specific information-the titles, authors, publishers, and ISBN/ASIN data-to plaintiffs' counsel for all downloads from the two pirated libraries. By August 15, 2025, parties must submit an agreed-upon notice for review and a plan to distribute notice to potential class members.

By its own calculation, Anthropic's statutory damages could be up to $150,000 times the five million books. Actual damages may be calculated in part by the price Anthropic later paid for print copies of the copyrighted works.

According to Judge Alsup, Anthropic "has actively submitted that piracy is for an AI company the fair price to pay."

The *Bartz v. Anthropic PBC* trial is scheduled to begin on December 1, 2025.