

Title: Anthropic’s Copyright Settlement: Lessons for AI Developers and Deployers

Category: Advanced Technology

Date” 9/17/25

Authors:

Harry A. Valetk	212 440 4416	harry.valetk@bipc.com
Patrick C. Keane	703 838 6522	patrick.keane@bipc.com
Lloyd Smith	202 452 7302	lloyd.smith@bipc.com

Summary: Anthropic, the developer of the Claude AI system, has agreed to a proposed \$1.5 billion settlement to resolve a class-action lawsuit, in which authors and publishers alleged that Anthropic used pirated copies of books to train its Large Language Models (LLMs).

Approximately 500,000 works are covered, with compensation set at approximately \$3,000 per book. As part of the settlement, Anthropic has also agreed to destroy the unlawfully obtained files.

Learn more about the lawsuit and its implications for companies developing AI tools.

Link: <https://www.bipc.com/anthropic%E2%80%99s-copyright-settlement-lessons-for-ai-developers-and-deployers>

Article: Anthropic, the developer of the *Claude AI* system, has agreed to a proposed \$1.5 billion settlement to resolve a class-action lawsuit, in which authors and publishers alleged that Anthropic used pirated copies of books — sourced from online repositories such as *Books3*, *LibGen*, and *Pirate Library Mirror* — to train its Large Language Models (LLMs).

Approximately 500,000 works are covered, with compensation set at approximately \$3,000 per book. As part of the settlement, Anthropic has also agreed to destroy the unlawfully obtained files.

Background

The case, *Bartz v. Anthropic* (3:24-cv-05417 (N.D. Cal.)), was filed in August 2024 by authors Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, and, from the

outset, the case presented a test of how traditional copyright doctrines would be applied to LLM training.

In June 2025, Judge William Alsup issued a pivotal summary judgment ruling in which he held that Anthropic's use of lawfully acquired books for AI training was "quintessentially transformative" and, thus, protected by fair use. However, he simultaneously ruled that Anthropic's creation and retention of a "central library," comprised in part of the pirated works, was not transformative and constituted infringement. This partial win-and-loss positioned the case for trial on damages.

The settlement is now subject to court approval, and Judge Alsup has already raised questions about adequacy and notice procedures. While final terms may evolve, the scale of the payout underscores the risks tied to sourcing training data from unauthorized repositories.

Compared to *Authors Guild v. Google* (2015)

This case is different from the landmark Google Books litigation, which the Second Circuit held that Google's mass digitization of books — used to create a searchable index with snippet displays — qualified as fair use. In that case:

- **Transformative Purpose:** Google's use enabled new research and indexing functionality, not substitution for the original works.
- **Scope of Use:** Full-text copying was allowed because it was necessary to enable search; access to entire texts was restricted.
- **Market Impact:** The court found no cognizable harm to the market for books.

By contrast, Anthropic's liability here stems not from the act of training *per se*, but from the unlawful acquisition of pirated materials. The distinction underscores that lawful sourcing of data is essential to mount a fair-use defense.

Recommendations for Companies

1. **Lawful Sourcing is Non-Negotiable.** Using pirated or unauthorized datasets exposes companies deploying AI Systems to extraordinary liability — even if the downstream use could be argued as transformative. The class certification ruling also suggests that aggregation of claims is viable, magnifying risk exposure. The *Bartz* litigation could also resonate beyond *Anthropic*. In fact, another group of plaintiffs involving music publishers have already moved to amend their complaint against *Anthropic* last month to add piracy claims. We may, therefore, see other AI platforms settle with creators as cases get closer to trial to avoid potentially astronomical damages. Companies using AI Systems should, therefore, negotiate contractual assurances or warranties that the AI System developer has conducted thorough reviews of its AI training inputs and has eliminated any reliance on questionable datasets such as *Books3* or other gray-market repositories.

2. Expect More Licensing Deals. The settlement and related regulatory actions (e.g., French fines against Google for AI use of publisher content) are accelerating a shift toward direct licensing arrangements with authors, publishers, and content platforms. Consider asking or including contractual assurances that the AI System developer has content licenses with publishers, authors, and collective rights organizations.

3. Settlements Do Not Clarify the Law. Anthropic's agreement avoids a definitive court ruling on fair use in the AI-training context. For now, *Authors Guild v. Google* remains the leading precedent, but its facts are distinguishable to most AI use cases. It is, therefore, good practice to track U.S. case law and EU regulatory enforcement activity, as both will shape the boundaries of permissible AI training practices. Companies should also maintain detailed records of dataset provenance to support compliance and defense strategies.

4. Regulatory and Litigation Risks Are Growing. US litigation is increasing, and EU regulators are scrutinizing AI firms' content practices. Companies should anticipate ongoing enforcement and lawsuits and consider incorporating copyright compliance and licensing obligations into enterprise AI governance and risk programs.

The Anthropic settlement underscores the financial and reputational risks of training AI systems on unlawfully obtained works. Companies developing or deploying AI tools should take proactive steps now to ensure that their data sourcing strategies withstand both judicial and regulatory scrutiny.

Click [here](#) to read the full summary judgment decision by Judge Alsup.