The case we are studying is *Bartz v. Anthropic*. The plaintiffs (Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson) made a class-action complaint against the defendant, Anthropic PBC illegally obtained their copyrighted works, as well as hundreds of thousands of other authors' books, to train their large language models (LLMs), and the company has made no attempt to compensate those authors. Without the works of those authors, Anthropic would never have been able to train Claude, but in using the pirated material, anyone can use Claude to generate a novel instantly and sell it. They are also able to create knockoffs of existing books, which further take away from authors' profits. This threatens the ability of authors to make a living off of their books, which take years to create. Therefore, they key stakeholders in this case are the three plaintiffs, all other authors and publishers whose work was or could have been used to train LLMs, AI developers, AI end-users that may use Claude or other AI models to generate creative work, and anyone else in creative fields (photographers, artists, musicians, etc.). Authors of textbooks and other academic publications could also be affected.

Anthropic obtained the books from "The Pile," which is "an 800 GB+ open source dataset created for large language model training." One dataset in The Pile, known as "Books3" contained "all of Bibliotik," a known pirating website containing just under 200,000 books. Anthropic used this dataset, in addition to data from other pirating websites, to train Claude, despite knowing that the sources were illegitimate. In the *Bartz v. Anthropic* case, the judge ruled that if the works had been legally obtained, using them to train the models may have constituted fair use. However, since the works were obtained illegally, Anthropic committed copyright infringement. Given this, the main problem is in the "Procurement & Vendor Selection" and "Data Practices" steps of the pipeline. Anthropic acquired its data from known pirating websites and datasets, and therefore, it did not have permission from the authors to use their work in that training. This directly violates authors' intellectual property rights and undermines their autonomy over their own creative labor.

Various other problems arise throughout other steps of the ethical pipeline. In the "Policy Goals & Success Metrics" step, the problem lies in the fact that Anthropic used pirated works to save time and money in launching Claude, so there was no consideration to compensating the authors and publishers properly. The "Workflow Use" step involved integrating illegally acquired books directly into training, with little documented review or risk-assessment processes. The speed amplified downstream harms, including the risk that the models trained on pirated works may reproduce copyrighted text. There is also a problem in the "Outcomes" step, because anyone can use Claude to create their own novels in minutes or hours. This brings financial harm to authors because their work has to compete with AI-generated novels, in addition to losing money from the illegal acquisition of their work to train the models in the first place. The benefits and burdens are not fairly distributed, because Anthropic and end-users generating creative work are the only ones that benefit, at the expense of all the authors and publishers, as well as anyone else in creative fields, who could experience a financial burden.

Works Cited

"Anthropic's Copyright Settlement: Lessons for AI Developers and Deployers." *Buchanan Ingersoll & Rooney PC*, 17 Sep. 2025, https://www.bipc.com/anthropic%E2%80%99s-copyright-settlement-lessons-for-ai-developers-and-deployers.

Bartz et al. V. Anthropic PBC, 2024, *ClassAction.org*, United States District court for the Northern of California, Case No. 3:24-cv-05417, https://www.classaction.org/media/bartz-et-al-v-anthropic-pbc_1.pdf.

AI Use Disclosure:

Tool(s) Used: Boodlebox, ChatGPT

What we used it for: Finding sources

Verified: Verified that it gave real sources related to the case that we are studying