

When machine learns from books: What Indian law must fix

India's first AI copyright case tests how far the law can stretch to keep up with machine learning, global precedents, and the idea of transformative use.



BY KALINDHI BHATIA

In the age of generative artificial intelligence, the tension between technological innovation and intellectual property protection is becoming increasingly pronounced. A recent lawsuit in the United States involving Anthropic PBC, a prominent AI company founded by former OpenAI employees, has brought this issue to the forefront.

The dispute centres around the use of copyrighted literary works to train large language models (LLMs) that power Anthropic's AI chatbot, Claude. This case, while rooted in American copyright law, offers important learnings for India, where similar disputes are beginning to emerge. As India's first AI copyright litigation unfolds in the ANI vs. OpenAI case, the legal frameworks governing AI training, particularly the differences between the American fair use doctrine and India's fair dealing exception, warrant closer examination.

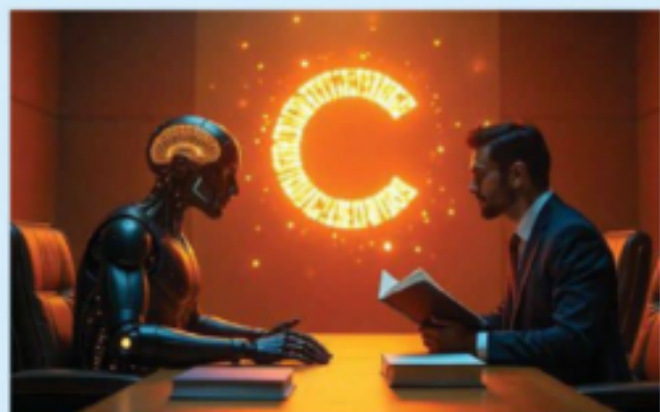
This legal battle began when authors Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson accused

Anthropic of unlawfully using their copyrighted works to train Claude. They alleged that Anthropic had not only pirated books from illegal sources such as Books3 and Library Genesis but also destructively scanned millions of lawfully purchased print books to create a massive internal digital library. This library served as a foundational dataset for training the Claude model. The authors claimed these acts constituted direct copyright infringement and undermined the emerging market for licensing books for AI training purposes.

ANTHROPIC'S DEFENCE: THE US FAIR USE CLAUSE

Anthropic mounted a strong defence under the US doctrine of fair use, which is codified in Section 107 of the Copyright Act of 1976. The company argued that training an LLM constitutes a transformative use—it does not simply copy or reproduce works but instead enables the model to generate new, original content. According to Anthropic, this process is analogous to how humans read books to learn style, tone, and structure.

India's fair-dealing law, unlike the US fair-use doctrine, leaves courts with limited tools to evaluate AI's transformative potential.



IN BRIEF

- Anthropic's US lawsuit highlights the global copyright challenge of training AI models with copyrighted literary data.
- The US recognises transformative use under fair use law, allowing certain types of AI training using copyrighted content.
- India's fair-dealing provision under Section 52 does not cover transformative or commercial use, limiting legal flexibility.
- ANI's lawsuit against OpenAI marks India's first major legal challenge to AI content reproduction and attribution.
- The Delhi High Court must now decide whether Indian law can meaningfully address global AI training and content rights.
- Without reform, India's copyright regime may fail to protect creators or support innovation in AI development.

Additionally, Anthropic claimed that the digitisation of lawfully purchased books for internal use was not an act of infringement, particularly because the physical copies were destroyed and the digital versions were not distributed publicly. The company also pointed out that its Claude model included filtering software to prevent the regurgitation of original texts, framing the AI's learning process as non-expressive and functional rather than creative replication.

US COURT RULING: WHAT IS FAIR, WHAT IS NOT

Unlike India's not-so-well-defined fair-dealing regime, American courts apply a flexible, four-factor test. These factors include: the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality of the portion used, and the effect of the use upon the market.

While interpreting Section 107 of the US Copyright Act of 1976, the court agreed with Anthropic on two crucial points. First, it ruled that the use of copyrighted material to train LLMs was transformative and constituted fair use. Second, the digitisation of lawfully purchased books, when used internally and without distribution, was also deemed permissible.

However, the court drew a clear line when it came to pirated books. The creation and indefinite retention of a digital library using content from unauthorised sources was found to be beyond the scope of fair use, especially given its scale and lack of defined purpose.

COPYRIGHT IN INDIA: THE FAIR DEALING DOCTRINE

India's approach to copyright law differs from that of the United States. Section 52 of the Indian Copyright Act, 1957 provides for a fair-dealing exception, but it is much narrower in scope. It applies only to specific purposes, such as private research, criticism, review, reporting of current events, and use in certain educational or archival contexts.

Transformative uses are not explicitly recognised within this framework. Indian courts have consistently interpreted the fair-dealing provision restrictively. In *Tekla Corporation vs. Survo Ghosh* case, the Delhi High Court emphasised that courts cannot extend the statutory exceptions through judicial interpretation.

TRANSFORMATIVE USE: A GAP IN INDIA'S LAW

A notable gap in Indian copyright law is the lack of an explicit exception for transformative use. Unlike the US, where courts assess whether a new work adds value or meaning to the original, Indian law does not formally recognise this concept. However, Indian courts have, in limited instances, applied similar reasoning. For example,

The ANI vs. OpenAI case may mark a constitutional moment for Indian copyright, forcing lawmakers to face the realities of machine learning.

In *Syndicate of the Press vs. BD Bhandari*, the Delhi High Court suggested that using a work for a “substantially different” purpose might not be infringement.

While this line of thinking resembles the US doctrine, it remains underdeveloped in Indian jurisprudence and has not been consistently applied in the context of modern technologies. Consequently, the law in India is currently ill-equipped to handle challenges relating to the use of copyrighted materials in AI training.

ANI VS. OPENAI: THE AI COPYRIGHT BATTLE

These challenges are being put to the test in what is arguably India’s first major AI copyright case: *Asian News International (ANI) vs. OpenAI*. Filed in the Delhi High Court in November 2024, ANI alleged that OpenAI’s chatbot, ChatGPT, generated responses that included unauthorised excerpts from ANI’s news articles. It also accused ChatGPT of fabricating news stories falsely attributed to ANI.

OpenAI defended its actions by citing the doctrines of fair use and transformative use, the principles rooted in US law. It also argued that since its servers are located outside India, the Delhi High Court lacked jurisdiction. To mitigate the dispute, OpenAI claimed it had already blocklisted ANI’s domain to prevent further training or output generation based on its content. However, these defences conflict with India’s legal framework.

The Delhi High Court has thus far focused on two pivotal questions: whether it has jurisdiction over OpenAI, a foreign entity, and whether the use of ANI’s content constitutes infringement under Indian copyright law. The appointment of two legal experts to assist the court highlights the complexity and significance of the matter. The case is ongoing, but its resolution could reshape how AI companies operate in India and how Indian copyright holders protect their works in the digital age.

TWO SYSTEMS: US VS. INDIAN FRAMEWORKS

A key takeaway from both the *Anthropic* case and the ANI litigation is the stark contrast between the US and Indian legal approaches to copyright. The US fair use doctrine provides courts with broad discretion and a flexible framework to assess new and evolving technologies. The

recognition of transformative use allows American courts to weigh public benefit, innovation, and market impact in a way that encourages technological progress.

In contrast, India’s fair-dealing regime is rigid. It restricts permissible uses to a specific list, leaving little room for judicial discretion or adaptation to digital realities. The lack of recognition for commercial or transformative uses further complicates matters for AI companies seeking to operate legally in India. Courts are left with limited tools to address nuanced questions relating to innovation and authorship.

DIGITAL ERA NEEDS: WHAT INDIA MUST FIX

India is quickly becoming a major player in the global AI landscape, with the domestic market projected to grow to USD 17 billion by 2027. This growth brings with it an urgent need to reform the country’s intellectual property regime to accommodate new realities. Historically, India has shown a willingness to align with international legal trends, as seen in the 2012 copyright amendments that incorporated provisions from the WIPO Internet Treaties.

As landmark global cases—*New York Times vs. OpenAI*, *Dow Jones vs. Perplexity*, and *Getty vs. Stability AI*—make their way through courts, their outcomes are poised to influence Indian jurisprudence. The ANI vs. OpenAI case, therefore, is more than just a legal contest. It marks a potential inflexion point in Indian copyright law, one that could prompt lawmakers and judges to re-examine the contours of fair dealing in a digital and AI-driven age.

To remain globally relevant and innovation-friendly, India’s legal framework must evolve. Courts and legislators will need to adopt a more adaptive approach—potentially recognising transformative use or expanding the interpretation of Section 52—to address the complex realities of AI-generated content. The future of AI in India will depend not only on domestic policy decisions but also on how thoughtfully the country engages with emerging global legal standards. 🌟

*The author is a Partner at BTG Advaya.
(with contribution from Urjaswal Bhatt).*

feedbackvnd@cybermedia.co.in

