What have courts ruled with respect to AI and copyright?

Can AI models be trained on datasets which contain both copyrighted and public domain content?

G.S. BAJPAI

The story so far:

Whether the intellectual material produced by various generative AI models infringes copyright laws has been a controversial question posed around the globe. Three recent rulings in the U.S. — *Thomson Reuters versus Ross Intelligence* (2025), *Bartz versus Anthropic* (2025), and *Kadrey versus Meta* (2025) — have brought considerable clarity to the issue. The decisions confirm that transformative training on legitimately acquired texts can qualify for 'fair use', though key limits remain on pirated content and unclear market impacts. However, the issue remains unresolved from a legal perspective.

Do AI models violate copyright law?

Generative AI models can occasionally produce content that closely resembles or even duplicates specific works from their training datasets, raising concerns about ethics and law. Legal outcomes often depend on whether training AI on original works and its subsequent output undermines the original works' market by replacing them, or whether the AI-generated content adds value and is considered transformative rather than a substitute. The legality of training AI with copyrighted data remains unsettled at the global level. Training generative AI models involves feeding them large datasets, often scraped from the internet, that include both copyrighted and public domain content, which raises legal issues regarding reproduction rights under copyright law. The primary concern is whether copying originals for training constitutes infringement or qualifies as fair use (in the U.S.) or as a text and data mining exception (in the EU and U.K).

What about databases?

The general principles of liability in determining the usage of databases and published works in the training of generative AI models are grounded in Intellectual Property (IP) law, contractual obligations, and privacy regulations. Generative AI has many IP uncertainties. There is legal ambiguity in determining whether the training of AI using IP-protected data, and the generated outputs constitute IP infringements. Some nations provide IP law exceptions, on the basis of it being for fair use, text and data mining, and temporary copying that may apply in cases involving generative AI. However, the absence of global harmonisation and the actual application of generative AI exceptions has not yet been tested, throwing up further legal uncertainty. Additionally, the ownership of IP rights of the output of generative AI is legally uncertain.

Presently, there is no explicit or harmonised global regulation that addresses the intellectual property implications of generative AI. The intellectual property laws of most nations were developed long before the advent of AI, leading to legal uncertainty over whether IP rights can subsist in AI-generated outputs and, if so, who would own them. This uncertainty is most pronounced in the area of copyright, where authorship traditionally requires human creativity.

What did the U.S. judgments state?

The two landmark U.S. court judgments, one in favour of Anthropic and the other Meta, deduce that the use of copyrighted material for training AI systems could qualify as fair use. However, these rulings do not close the debate regarding the legality of sourcing training data from pirated repositories.

In the *Anthropic case*, Judge William Alsup of the District Court in the Northern District of California ruled that using copyrighted data for training AI software was transformative, comparing the model's training to a writer learning from prior works. However, the judge held that Anthropic must face trial over its use of pirated copies to develop its library of material.

In the *Meta* case, Judge Vince Chhabria of the Northern District of California ruled in Meta's favour, concluding that the plaintiffs had not established that the company's use of their works would result in market dilution by generating AI outputs like the originals. Meta's actions were considered to be covered under the 'fair use' provision. But the judge said that tech companies making money off the AI boom ought to figure out ways to share the wealth with companies that hold copyrights. In both rulings, the judges adopted a broad view of the concept of 'fair use' when applied to

AI training, and provided tech firms with legal protection from copyright liability. But the concerns of unauthorised data harvesting, or of future market damage, have not been dealt with. Courts have signalled that piracy is still a liability and that compensation systems for creators are long overdue.

What are the implications for India?

The *ANI versus OpenAI* lawsuit is significant in clarifying how India's existing IPR framework applies to generative AI. Under the Copyright Act, 1957, copyright owners enjoy exclusive economic rights including reproduction, adaptation, and translation, which require permissions for commercial use unless an exception under Section 52 (fair dealing) applies. While some argue that India's IP laws lack provisions specific to AI, the official position holds that the current legal framework is sufficient to address AI-related issues. India, as a member of major international IP treaties, recognises works created by legal persons and provides mechanisms to enforce rights through both civil and criminal remedies, including measures against digital circumvention.

The author is Vice Chancellor, National law University Delhi