



Who Owns Internet's Knowledge? Copyright, AI Training and Indian Law

News & Updates • October 20, 2026

Every time someone publishes an article, uploads a photograph, releases a song or shares research online, they contribute to an expanding body of human knowledge. Generative AI systems can draw on enormous quantities of digital material to learn patterns, generate responses and perform increasingly complex tasks.

But accessibility is not the same as permission. If a machine learns from millions of copyrighted works, who gets to decide how those works are used? And when that learning becomes the foundation of a commercial AI product, should the original creators be entitled to compensation?

These questions are no longer theoretical in India. They sit at the intersection of copyright law, technological innovation and the economic value of human creativity.

When AI Learns From Copyrighted Works

Training a large language model involves processing extensive datasets. These may contain books, news reports, articles, images and other material protected by copyright. The legal concern is not simply that an AI system learns from existing information. It is whether the process involves acts restricted by copyright law, including reproducing protected expression without authorisation.

India's Copyright Act, 1957, grants copyright owners specified exclusive rights under Section 14. Section 52 provides exceptions for certain uses, including fair dealing for private or personal use, including research. However, Indian copyright legislation does not contain an express, general text-and-data-mining exception specifically designed for AI training.

This leaves a difficult question: can existing exceptions accommodate a technology that processes copyrighted works at a scale and for purposes that the law did not expressly contemplate?

What the ANI v. OpenAI Dispute Reveals

The issue came into sharp focus in *ANI Media Pvt. Ltd. v. OpenAI OpCo LLC*, before the Delhi High Court.

ANI alleged that OpenAI had used its copyrighted news articles to train ChatGPT without authorisation. It also raised concerns about outputs allegedly reproducing its reporting. OpenAI



disputed the claims, including the allegation that its training activities infringed ANI's copyright.

On 24 July 2026, the Delhi High Court declined to grant ANI an interim injunction. At this preliminary stage, the Court found that the challenged use of ANI's works for model training fell within the research exception under Section 52(1)(a). Its assessment considered the purpose of the use, potential harm to ANI's legitimate commercial interests and the public interest in technological research.

The Court also examined the separate allegations concerning ChatGPT's outputs. It distinguished model training from retrieval-augmented generation, a process through which a system retrieves external information when responding to a prompt. On the material before it, the Court found no sufficient basis for an interim finding that the challenged outputs substantially reproduced ANI's protected expression.

The distinction matters. The use of copyrighted material to train a model and the reproduction of copyrighted expression in an output are separate legal questions. A conclusion concerning one does not automatically resolve the other.

The order is significant, but its findings are preliminary. The litigation remains pending, and the decision does not establish blanket permission for all AI developers to use copyrighted material without a licence.

Innovation Versus the Creator's Right to Control

AI developers have a legitimate interest in building systems that advance research, accessibility and productivity. Authors, publishers, artists and other rights holders have an equally important interest in protecting the value of their work.

The difficulty lies in balancing these interests without treating either technological progress or copyright ownership as absolute.

A model trained on copyrighted works may produce a new response rather than reproduce a source passage verbatim. Yet the absence of verbatim reproduction in an output does not, by itself, answer whether unauthorised copying occurred during training. Conversely, the fact that a work was included in training data does not automatically establish infringement without examining the relevant rights, acts and exceptions.

The legal analysis must therefore distinguish the source material, the act of copying, the purpose of use and the resulting output.



Can Licensing Provide a Middle Ground?

In December 2025, the Department for Promotion of Industry and Internal Trade (DPIIT) published the first part of a [working paper on the AI-copyright interface](#). It proposed a hybrid framework under which AI developers could use lawfully accessed copyrighted material for training under a blanket licence, with royalties payable upon commercialisation and administered through a centralised mechanism.

The proposal seeks to reduce the burden of negotiating individual licences while recognising creators' economic interests. It remains a policy proposal, not an enacted amendment to the Copyright Act.

Its significance lies in the question it raises: must creators choose between protecting their work and allowing innovation, or can a workable licensing framework support both?

Who Benefits From the Knowledge AI Uses?

The internet's knowledge has been built through countless acts of human creativity. AI can make that knowledge more accessible and useful, but the legal framework must still address how protected works enter these systems and how their value is distributed.

India's evolving jurisprudence will help determine where research exceptions end, when licensing is necessary and how copyright law should respond to technologies that transform the use of creative material.

The challenge is not simply to decide whether AI may learn from human work. It is to determine the conditions under which that learning can occur without making the rights of the people who created the underlying material irrelevant.



RELATED INDUSTRIES

[ARTIFICIAL INTELLIGENCE](#)

RELATED PRACTICES

[COPYRIGHT](#)