Can AI Training Be Copyright Infringement? The ANI v. OpenAI Dispute in India

News & Updates • November 14, 2026

Generative artificial intelligence systems are trained on large volumes of information, including text, images, news articles and other digital material. Some of this material may be protected by copyright. This raises a fundamental question: when an AI developer collects and processes copyrighted works to train a model, does that activity constitute copyright infringement?

The question has acquired particular significance in India through *ANI Media Pvt. Ltd. v. OpenAI OpCo LLC*, a copyright dispute before the Delhi High Court. The proceedings bring the relationship between copyright ownership, machine learning and the statutory exceptions under the Copyright Act, 1957 into focus.

What Is the Dispute About?

In November 2024, news agency Asian News International (ANI) initiated proceedings against OpenAI, alleging unauthorised use of its copyrighted news content to train the large language models underlying ChatGPT.

ANI's claims concerned two principal activities. First, it alleged that its copyrighted articles were collected, copied and stored for AI training without authorisation. Second, it contended that ChatGPT generated responses reproducing or drawing upon its copyrighted material, including content that was allegedly presented as ANI reporting.

OpenAI disputed these allegations. Among other arguments, it maintained that its models learn linguistic patterns and information rather than retain and reproduce copyrighted works as such. It also relied on the research exception under Section 52 of the Copyright Act.

The proceedings therefore raised questions about both the reproduction of copyrighted material during model training and the nature of content generated in response to user prompts.

Copyright Protection Under Indian Law

Section 14 of the Copyright Act, 1957 defines the exclusive rights available to copyright owners. For literary works, these include the right to reproduce the work in material form, issue copies to the public and communicate the work to the public, among other rights.

Under Section 51, unauthorised exercise of rights reserved for the copyright owner may constitute infringement, subject to the statutory exceptions.

One such exception appears under Section 52(1)(a), which addresses fair dealing with a work for specified purposes, including private or personal use, research, criticism or review, and reporting of current events and current affairs.

Unlike certain jurisdictions that have introduced specific text-and-data-mining exceptions, India's Copyright Act does not expressly provide a standalone exception for AI training. The scope of the existing fair-dealing provision has consequently become central to the dispute.

What Did the Delhi High Court Decide?

On 24 July 2026, the Delhi High Court, in its interim order, declined ANI's application seeking an injunction against OpenAI.

The Court examined whether storing ANI's copyrighted works for training large language models amounted to infringement and whether the use fell within Section 52(1)(a).

At the prima facie stage, the Court considered the purpose of the training activity, the nature of the use and the broader research objectives associated with developing large language models. It concluded that OpenAI's storage of ANI's works for model training fell within the research exception under Section 52(1)(a) and therefore did not amount to infringement at that stage.

The Court also considered the separate question of AI-generated outputs. It found that the responses produced through retrieval-augmented generation (RAG) were not substantially similar to ANI's original works. It further observed that ANI had not established memorisation and regurgitation of its copyrighted material through ChatGPT's responses.

Importantly, the Court's observations were expressly confined to the interim application and were not to determine the final outcome of the suit.

Why Does the Distinction Between Training and Outputs Matter?

AI training and AI-generated outputs involve different copyright questions.

AI developers may collect, reproduce and process copyrighted material to train a model. The legal question concerns whether those activities engage the exclusive rights of copyright owners and, if so, whether an applicable statutory exception covers them.

Outputs raise a separate question: whether the material generated by an AI system reproduces a substantial part of a protected work. The use of facts or general information from a news article does not necessarily amount to reproduction of its protected expression.

Retrieval-augmented generation adds another dimension. Unlike model training, RAG retrieves external information when responding to a prompt. The legal assessment may therefore depend on the source material accessed and the content ultimately reproduced in the response.

These distinctions are relevant because the legality of training does not automatically determine whether every AI-generated output is lawful.

The Appeal and the Continuing Legal Questions

ANI challenged the July 2026 order before a Division Bench of the Delhi High Court. On 15 September 2026, the Division Bench issued notice to OpenAI and declined to grant immediate interim relief. The Division Bench scheduled the appeal for further hearing on 8 December 2026. The dispute remains significant because it requires courts to interpret existing copyright provisions in light of technologies that legislators did not specifically contemplate when they enacted the law.

The case also raises wider questions concerning the reproduction of copyrighted works during machine learning, the scope of research under Section 52, the treatment of AI-generated outputs and the territorial application of Indian copyright law where training infrastructure is located outside India.

The ANI proceedings do not establish a general rule that all AI training using copyrighted material is permissible. Equally, the use of copyrighted material in training does not, by itself, resolve every question of infringement. The legal assessment remains connected to the rights asserted, the nature of the activity and the applicability of statutory exceptions.

As the appeal and the underlying suit progress, the interpretation of the Copyright Act, 1957 in relation to generative AI will remain an important area of Indian intellectual property law.

RELATED INDUSTRIES

ARTIFICIAL INTELLIGENCE

RELATED PRACTICES

COPYRIGHT