
WHEN THE MACHINE READS THE LIBRARY: ANI V. OPENAI AND THE QUESTION SECTION 52 CANNOT ANSWER

Pranav Ankush Jadhav, B.A. LL.B., ILS Law College, Pune

ABSTRACT

The pending Delhi High Court case of Asian News International (ANI) v OpenAI Inc raises a deceptively simple question: does training a large language model on copyrighted news content constitute infringement under the Copyright Act, 1957? The answer turns out to be anything but simple. Section 52 of the Act sets out an exhaustive list of permitted acts, and neither the fair dealing provision under Section 52(1)(a) nor the electronic storage exception under Section 52(1)(n) maps cleanly onto what AI companies do when they scrape and process published works at scale. This article examines the doctrinal split that emerged from the submissions of the two amici curiae appointed by the court, analyses why both positions leave the output-stage reproduction question unanswered, and argues that the structural limits of Section 52 cannot be resolved by judicial interpretation alone. It further contends that India needs a narrow, opt-out text and data mining exception modelled on the EU's CDSM Directive — coupled with explicit output-stage liability — if copyright law is to remain coherent in the age of generative AI.

Introduction

The Copyright Act, 1957 was written for human beings. A student who reads a hundred books to prepare for an exam does not infringe any of them. A journalist who reviews archived reports before filing a story, a researcher who works through a body of published articles, none of them need a licence. Reading, even exhaustive reading, is invisible to copyright law.

Now consider a large language model. It processes millions of news articles. It extracts statistical patterns from them. It uses those patterns to generate text that users ask for. Is that reading? Or is it something the Copyright Act never contemplated?

This question is before the Delhi High Court in *Asian News International (ANI) v OpenAI Inc.*,¹ the most consequential copyright case in India right now, currently at an interim stage before the Delhi High Court — not because of its facts, which are straightforward enough, but because of what it reveals about the limits of Section 52.

I. The Facts

ANI, one of India's largest news wire services, alleges that OpenAI scraped its copyrighted news content to train ChatGPT without a licence. It further alleges that ChatGPT reproduces ANI articles verbatim in user responses. OpenAI's defence is that training constitutes 'non-expressive use' the model extracts patterns, not content and that any incidental reproduction is transformative.

The court appointed two amici curiae: Prof. Arul George Scaria, a copyright scholar at NLU Delhi, and Mr. Adarsh Ramanujan, Senior Advocate. Their submissions produced a doctrinal split the court has not yet resolved. That split exposes exactly where Indian copyright law is unprepared.

II. What Section 52 Says — and What It Does Not

Section 52 of the Copyright Act lists acts that do not constitute infringement. The list is exhaustive. The Supreme Court confirmed this in *Eastern Book Co v DB Modak*,² and the Delhi High Court reinforced it in *Super Cassettes Industries v Chintamani Rao*,³ holding that courts cannot devise new exceptions to accommodate technological change. Whatever is not in the

¹*Asian News Int'l (ANI) v. OpenAI Inc.*, No. CS(COMM) 428/2024 (Delhi H.C. filed 2024).

²*E. Book Co. v. D.B. Modak*, (2008) 1 SCC 1.

³*Super Cassettes Indus. Ltd. v. Chintamani Rao*, 2012 SCC OnLine Del 2901 (Delhi H.C. 2012).

list is simply not protected.

The two provisions most relevant to this dispute are Section 52(1)(a)⁴ — fair dealing for private or personal use including research, criticism, review, and reporting of current events and Section 52(1)(n),⁵ which permits electronic storage for preservation purposes.

Neither fits AI training cleanly. The 'private or personal use' limb has consistently been read by Indian courts as covering human, individual, non-commercial activity.⁶ OpenAI is not a student studying news. It is a commercial entity building a product that competes with the publishers it scraped. The 'research' limb is harder to dismiss, and this is exactly where the amici diverge.

III. The Amicus Split — and Where Each Argument Breaks Down

Prof. Scaria argued that machine learning is not categorically different from human learning. A model that reads text, extracts patterns, and retains statistical information is doing what any researcher does. On this view, if training does not make expressive content publicly available, that is, the original article is neither stored nor served to users — it falls within Section 52(1)(a).

Mr. Ramanujan took the opposite position. Commercial copying at scale, for the purpose of building a product, is not 'research' in any sense Section 52 recognises. The purpose is product development. Fair dealing was never meant to absorb that cost on behalf of commercial enterprises.

Both have force. Scaria is right that the training stage, viewed in isolation, may not produce infringement in the traditional sense, no article is distributed or commercially exploited. Ramanujan is right that purpose controls the analysis, and the purpose here is unambiguously commercial. It may also be argued that AI outputs are probabilistic and do not store or retrieve exact copies of copyrighted works, making verbatim reproduction relatively rare. However, where such reproduction does occur and substitutes for the original work in the market, that defence becomes difficult to sustain.

But here is the problem with Scaria's position that no published commentary has yet addressed squarely: even if the training stage is non-expressive, that shield does not extend to OpenAI's

⁴Copyright Act, No. 14 of 1957, § 52(1)(a) (India).

⁵Copyright Act, No. 14 of 1957, § 52(1)(n) (India).

⁶*Civic Chandran v. Ammini Amma*, 1996 SCC OnLine Ker 149 (Kerala H.C. 1996) (interpreting the purpose-driven scope of fair dealing under § 52).

outputs. ANI's complaint is not only that OpenAI trained on its content. It is that ChatGPT reproduces ANI articles near-verbatim when users ask about news events. A user who receives that reproduction has received a substitute for the original. The market harm is direct. No provision in Section 52 covers this.⁷

So even accepting Scaria's more generous reading at the training stage, the case does not go away. The output stage is where OpenAI's exposure is clearest and where the law, as it stands, appears to leave little room for ambiguity.

IV. What the US and EU Have That India Does Not

Section 107 of the US Copyright Act⁸ provides an open-ended fair use defence governed by four factors: purpose and character of the use, nature of the work, amount copied, and market effect. US courts applied this to permit the Google Books project⁹ and the HathiTrust digital library.¹⁰ The framework allows a court to weigh all relevant considerations. India's Section 52 does not — it requires locating the use within one of the listed purposes, and if it falls outside, the analysis ends there. Indian courts have traditionally interpreted fair dealing more narrowly than the open-ended fair use doctrine in the United States, and this structural difference matters considerably when assessing AI training.

The EU went further. The CDSM Directive¹¹ introduced explicit text and data mining exceptions in Articles 3 and 4. Research organisations can mine freely. Commercial entities can mine lawfully accessed content, but rights holders can opt out by machine-readable declaration, preserving their ability to negotiate licences.

India has neither. A court interpreting Section 52 cannot import either approach. The statute is a closed list. This is not a gap that judicial interpretation alone can easily fill.

V. What Parliament Must Do

In early 2026, the Ministry of Commerce constituted an expert panel to assess whether India's copyright framework is adequate for AI.¹² The panel's recommendations should include three

⁷Copyright Act, No. 14 of 1957, § 14(a)(i) (India) (granting the author the exclusive right to reproduce the work in any material form; verbatim reproduction in AI outputs engages this right directly).

⁸17 U.S.C. § 107 (2022).

⁹*Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015).

¹⁰*Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).

¹¹Council Directive 2019/790, arts. 3–4, 2019 O.J. (L 130) 92 (EU).

¹²Ministry of Commerce & Indus., Gov't of India, Expert Committee on Artificial Intelligence and Copyright (2026).

things.

First, permit reproduction of lawfully accessed copyrighted works for text and data mining, including AI training. This is the basic permission the sector needs, and it should be limited to works the entity has lawful access to, not a blanket licence to scrape everything.

Second, give rights holders the ability to opt out of commercial TDM by machine-readable declaration. Publishers who want to license their content commercially can block free scraping. Publishers who do not object need do nothing. This restores the negotiating position that ANI and others should have had.

Third, and this is what neither the US nor the EU has adequately handled, create output-stage liability where AI responses are substantially similar to a specific copyrighted work in a way that substitutes for it in the market. Training and output must be treated as separate questions. Even if training is permitted under a TDM exception, verbatim reproduction in outputs should remain actionable. Without this, the exception becomes a subsidy for AI companies to train on content and then compete with the publishers they trained on. That is not a balance.

Conclusion

The Delhi High Court will decide *ANI v OpenAI* on the law as it stands. Given Section 52's structure, OpenAI's position at the output stage is difficult to defend. At the training stage the court may find room, but even then, the output issue does not disappear.

What no court can do is write a TDM exception into a closed statutory list. That is Parliament's job. A narrow, opt-out TDM exception with output-stage liability is not a radical proposal. It is the minimum update a 1957 statute needs to handle a 2024 technology.

The expert panel has a real chance to recommend it. The longer Parliament waits, the more India's AI sector operates on legal uncertainty, and the more India's publishers operate without protection. Both are bad outcomes, and both are avoidable.