



The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2026

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International (CC-BY-NC-SA 4.0) License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

Copyright vs. Artificial Intelligence: Examining the legal Challenges of AI Training and Authorship Under the Indian Law

- *IFFAT ZEHR*

ABSTRACT:

The rapid diffusion of generative artificial intelligence has unsettled two foundational assumptions of Indian copyright law that a work must trace back to a human author, and that the material used to build a creative system was lawfully obtained. This article examines the statutory architecture of the Copyright act, 1957, particularly the definition of author under Section 2(d) and its 1994 extension to computer generated works against the backdrop of judicial reasoning on originality developed in cases such as *Eastern Book Company v D.B. Modak* and *R.G. Anand v Delux Films*. It then turns to the more urgent and commercially consequential question of whether training large language models on copyrighted text amounts to infringement, using the Delhi High Court's 2026 order in *ANI Media Pvt Ltd v Open AI Inc* as a lens. The article argues that Indian law is currently ill-equipped on both fronts to deal with machine generated creativity and machine consumed data, and it proposes a set of legislative interventions drawing on comparative practice in the United Kingdom, the United States and Singapore.

Keywords: copyright, artificial intelligence, authorship, copyright Act 1957, fair dealing, text and data mining, ANI V OpenAI

(A) INTRODUCTION:

Ask a copyright lawyer in Delhi or Mumbai what an “author” is and the answer will almost always begin with a person. That instinct is not accidental; it is written into the structure of the Copyright Act, 1957, which has always assumed a human being sits somewhere behind every literary, artistic or musical work. Generative AI systems complicate that assumption in two distinct ways. First, tools such as large language models and diffusion based image generators

can now produce output with minimal, sometimes negligible, human direction raising the question of who if anyone should be recognised as the author of what merges. Second and arguably more pressing for Indian industry right now, these systems are trained on enormous corpora of pre-existing text, images and code much of it copyrighted without the consent of the original rights holder. The first problem is largely theoretical for now, the second has already reached the Delhi High Court. This article takes problems in turn, situates them within the existing statutory text and ask what if anything Parliament ought to do about either.

(B) AUTHORSHIP AND ORIGINALITY UNDER THE COPYRIGHT ACT, 1957:

The author of a literary or dramatic work is simply "the author," the author of a musical work is the composer, and the author of an artistic work other than a photograph is the artist (Copyright Act 1957, s. 2(d)). The definition of "author" varies depending on the type of work. Section 2(d)(vi), which was added by amendment in 1994, is more pertinent here. It states that the author of a computer-generated literary, dramatic, musical, or artistic work is "the person who causes the work to be created" (Copyright Act 1957, s. 2(d)(vi)). The clause was written when "computer-generated" meant something more akin to a spreadsheet macro than a neural network that could make its own stylistic decisions, and it demonstrates this: the clause assumes a human "cause" that is close enough to the machine to be recognized, an assumption that becomes more dubious when a model is able to generate a poem or a logo from a two-line prompt without additional human input (Rana and Joy, 2019).

Section 13, which limits protection to "original" works, governs originality in addition to authorship (Copyright Act 1957, s. 13). Since the Act does not define "original," it is up to the courts to do so. In *Macmillan & Co Ltd v. K and J Cooper* (*Macmillan v. Cooper*, 1924), the Privy Council's early formulation linked originality to "labour, skill and capital" (a low bar retained from English "sweat of the brow" notion). In *Eastern Book Company v. D.B. Modak*, the Supreme Court clarified this, rejecting both that minimal standard and the more stringent American "modicum of creativity" test in favor of requiring a work to exhibit "skill and judgment" that is not "so trivial that it could be characterized as a purely mechanical exercise" (*Eastern Book Co v D.B. Modak*, 2008). The Court had previously emphasized in *R.G. Anand v. Delux Films* that copyright protects expression rather than the underlying idea. This

distinction becomes problematic when a model's output is intentionally a statistical recombination of patterns from earlier works (*R.G. Anand v. Delux Films*, 1978).

(C) CAN A MACHINE BE AN AUTHOR? THE LIMITS OF SECTION 2(D)(VI):

Indian judges have been hesitant to extend authorship to non-human actors. The Delhi High Court ruled in *Rupendra Kashyap v. Jiwan Publishing House Pvt Ltd* that the Central Board of Secondary Education, an artificial legal body, could not be regarded as the "author" of exam papers because authorship under the Act is linked to a distinct creative agency (*Rupendra Kashyap v. Jiwan Publishing House*, 1994). An autonomous software system is much more vulnerable if a statutory body finds it difficult to be considered an author. The practical result of Section 2(d)(vi) is that copyright in AI-assisted output vests in whichever human "causes" the work, provided that causal link can be demonstrated at all (Ipleaders, 2023). No Indian statute or precedent authorizes an AI system itself to possess copyright. Current doctrine creates a real gap when a model creates something without a meaningful human curation of particular expressive choices: the work may meet the low originality standard imposed in *Eastern Book Company*, but it does not have an author recognized by the legislation.

This problem is not exclusive to India. India's 1994 amendment seems to be modelled after the United Kingdom's Copyright, Designs and Patents Act 1988, which grants authorship of a computer-generated work to "the person by whom the arrangements necessary for the creation of the work are undertaken" (Copyright, Designs and Patents Act 1988, s. 9(3)). The United States has taken a different approach. In *Thaler v. Perlmutter* (2023), a federal court upheld the Copyright Office's refusal to register a work created solely by an AI system without a human author, reiterating human authorship as a statutory prerequisite. This principle was first hinted at in the "monkey selfie" case, *Naruto v. Slater* (*Naruto v. Slater*, 2018). Since companies investing in AI-assisted content cannot be confident what, if anything, they own, the lack of an analogous declaration in India is a cost to the industry.

(D) AI TRAINING, FAIR DEALING AND THE ANI V OPENAI LITIGATION:

The second, more pressing issue pertains to the input side: is it legal to scrape and consume copyrighted content in order to train a model? Certain uses of copyrighted works, such as private study, research, criticism, review, and reporting of current events, are permitted by India's fair dealing clause, Section 52 (Copyright Act 1957, s. 52). In contrast to the UK, Singapore, and Japan, India does not have a specific statutory exception for text and data mining. As a result, developers must uncomfortably claim that using millions of copyrighted articles for training qualifies as "research" under Section 52(1)(a) (Ippress, 2024).

In the November 2024 Delhi High Court case ANI Media Pvt Ltd v. Open AI Inc., Asian News International claimed that OpenAI had trained ChatGPT on its news reports without a license and that the chatbot had created content that was falsely attributed to the agency (ANI Media v. OpenAI, 2024). OpenAI's defense mostly relied on the argument that training takes place on servers outside of India and, on the other hand, that training is a transformative process that publishers could theoretically prevent by crawler limits if they objected (Inc42, 2026). Relying on *Neetu Singh v. Telegram FZ* to support the claim that "conventional concepts of territoriality cannot be strictly applied" in the era of cloud computing (Delhi High Court, 2026a), the Court first determined that it had jurisdiction and rejected the notion that violators could move servers to avoid Indian law. On the merits of the interim application, however, the Court declined to grant an injunction, concluding that since copyright protects expression rather than the facts underlying news reports, ANI had not established a *prima facie* case that ChatGPT's outputs substantially reproduced its protected expression (World Trademark Review, 2026). The order marks India's first significant judicial involvement with LLM training, and its jurisdictional reasoning alone will influence future claims against international AI developers, even though the main litigation is still pending.

(E) TOWARDS A COHERENT FRAMEWORK:

It appears that three initiatives are long overdue. Instead of allowing developers to take refuge under an inappropriate "research" clause that was never intended for this purpose, Parliament should first consider a narrowly tailored text and data mining exception, along UK and Singaporean lines, that distinguishes non-commercial research mining from the wholesale commercial ingestion practiced by frontier AI labs. Second, where human participation is minimal, Section 2(d)(vi) has to be updated to specify who owns the output: the model

developer, the deploying entity, or, as some scholars suggest, a time-limited neighbouring right separate from full authorship (Rana and Joy, 2019). Third, publishers could be paid for the use of training data through a licensing or collective-management system that is akin to statutory licensing under Section 31D. This would eliminate the need for case-by-case litigation, which is cumbersome and can take years to reach a final merits ruling, as the ANI proceedings demonstrate.

CONCLUSION:

It is evident at both ends of the AI pipeline that Indian copyright law was not designed with autonomous machine innovation in mind. Regarding output, the 1994-era language of Section 2(d)(vi) tries to address the question of who, if anyone, is the creator of a work that a model generates primarily by itself. On the input side, courts must fit a truly innovative commercial technique into fair dealing categories designed for newsrooms and classrooms rather than server farms due to the lack of a text and data mining exception. The way the Delhi High Court handled ANI v. OpenAI demonstrates that while the judiciary is prepared to exercise jurisdiction over overseas AI researchers, it is nevertheless hesitant to declare infringement in the absence of convincing evidence of substantial replication. The way AI companies develop and license their models for the Indian market in the years to come will depend on whether that caution solidifies into established doctrine or shifts to a more protective posture once the main litigation is determined on merits.

REFERENCES:

1. ANI Media Pvt Ltd v Open AI Inc & Anr, CS(COMM) 1028/2024, Delhi High Court (India), filed 19 November 2024.
2. Copyright Act 1957 (India), No. 14 of 1957.
3. Copyright, Designs and Patents Act 1988 (UK), c. 48.
4. Delhi High Court (2026a) ANI Media Pvt Ltd v Open AI Inc & Anr, order dated 24 July 2026, discussed in Anand and Anand commentary, available at: <https://beta3.law.asia/?p=541594> (Last accessed on 20 September 2026).
5. Eastern Book Company v D.B. Modak (2008) 1 SCC 1 (Supreme Court of India).
6. Inc42 (2026) 'Delhi HC Rejects ANI's Plea for Interim Injunction Against OpenAI In Copyright Suit', Inc42, available at: <https://inc42.com/?p=565765> (Last accessed on 20 September 2026).

7. Ippress (2024) 'Copyright and AI: Deciphering Ownership and Authorship in AI-Generated Works', The IP Press, available at: <https://www.theippress.com/2024/02/04/copyright-and-ai-deciphering-ownership-and-authorship-in-ai-generated-works/> (Last accessed on 20 September 2026).
8. Macmillan & Co Ltd v K and J Cooper AIR 1924 PC 75 (Privy Council).
9. Naruto v Slater, 888 F.3d 418 (9th Cir. 2018) (United States Court of Appeals, Ninth Circuit).
10. Neetu Singh & Anr v Telegram FZ & Ors, Delhi High Court (India), 2022.
11. Rana, L. and Joy, M.M. (2019) 'Artificial Intelligence and Copyright – The Authorship', Mondaq / SS Rana & Co, 5 December, available at: <https://ssrana.in/?p=8144> (Last accessed on 20 September 2026).
12. R.G. Anand v Delux Films AIR 1978 SC 1613 (Supreme Court of India).
13. Rupendra Kashyap v Jivan Publishing House Pvt Ltd, 1994 PTC (16) 439 (Delhi High Court).
14. Thaler v Perlmutter, 687 F. Supp. 3d 140 (D.D.C. 2023) (United States District Court, District of Columbia).
15. World Trademark Review (2026) 'Delhi High Court Addresses Copyright and LLM Training in ANI v OpenAI', WTR, available at: <https://www.worldtrademarkreview.com/article/delhi-high-court-addresses-copyright-and-llm-training-in-ani-v-openai> (Last accessed on 20 September 2026).