

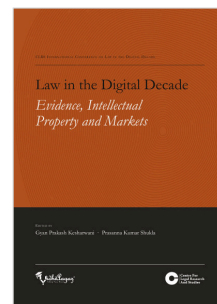
CHAPTER 9

Intellectual Property in the Age of Generative Technology: Rethinking Authorship, Training Data, and Infringement Liability

Shyamal M. Dave and Aditi Patel

in *Law in the Digital Decade: Evidence, Intellectual Property and Markets*, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026), pp. 85–91

DOI <https://doi.org/10.63108/VAB.LDD.2.9>



CITATIONS

BLUEBOOK (21ST ED.)

Shyamal M. Dave & Aditi Patel, *Intellectual Property in the Age of Generative Technology: Rethinking Authorship, Training Data, and Infringement Liability*, in *LAW IN THE DIGITAL DECADE: EVIDENCE, INTELLECTUAL PROPERTY AND MARKETS* 85 (Gyan Prakash Kesharwani & Prasanna Kumar Shukla eds., VidhiAagaz 2026).

OSCOLA (4TH ED.)

Shyamal M. Dave and Aditi Patel, 'Intellectual Property in the Age of Generative Technology: Rethinking Authorship, Training Data, and Infringement Liability' in Gyan Prakash Kesharwani and Prasanna Kumar Shukla (eds), *Law in the Digital Decade: Evidence, Intellectual Property and Markets* (VidhiAagaz 2026) 85.

CHICAGO (17TH ED.)

Dave, Shyamal M., and Aditi Patel. "Intellectual Property in the Age of Generative Technology: Rethinking Authorship, Training Data, and Infringement Liability." In *Law in the Digital Decade: Evidence, Intellectual Property and Markets*, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla, 85–91. Maihar: VidhiAagaz, 2026. <https://doi.org/10.63108/VAB.LDD.2.9>.

APA (7TH ED.)

Dave, S. M., & Patel, A. (2026). Intellectual property in the age of generative technology: Rethinking authorship, training data, and infringement liability. In G. P. Kesharwani & P. K. Shukla (Eds.), *Law in the digital decade: Evidence, intellectual property and markets* (pp. 85–91). VidhiAagaz. <https://doi.org/10.63108/VAB.LDD.2.9>

SOURCE

<https://doi.org/10.63108/VAB.LDD.2.9>

LICENCE

Open access under the Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0): it may be copied, shared and adapted for non-commercial purposes, with credit to the authors and the source.

Citations are given as a guide; check them against the style manual you follow. This page was added by VidhiAagaz to the downloaded file. The chapter follows, with the page numbers of the printed book.

CHAPTER 9

Intellectual Property in the Age of Generative Technology: Rethinking Authorship, Training Data, and Infringement Liability

DR. SHYAMAL M. DAVE* AND ADITI PATEL**

Generative AI systems are trained on very large bodies of copyrighted text, images and code, and whichever AI system is used, e.g. ChatGPT or DALL-E, raises another copyright question: who, if anyone, owns what the system produces? This paper investigates both extremes of this issue. As for inputs, it compares the methods of ingesting copyrighted works in training generative models: how the United States courts are starting to split in their decision and how, without a text-and-data-mining exception, Indian law remains significantly less clear than, for example, the EU's opt-out model under the AI Act. On the output side, it considers the requirement of human-authorship, now affirmed at the highest level of the U.S. courts to date, and whether this requirement is inescapably dictated, or simply permitted, by an originality standard under the Indian copyright law. The core of the paper is that India is at a genuine tipping point: It was not bound by any precedents on either question, there was an ongoing case before the Delhi High Court, and there was a government paper proposing a compulsory blanket-licensing system that had not yet been passed. The paper shares some lessons from the recent developments in the case law of the United States, Europe's regulatory framework, and India's shifting policy stance, and suggests that a hybrid model between full (or semi-) statutory licensing and reliance on fair-dealing litigation is more desirable than leaving the issue for the courts. It ends with concrete suggestions on how the Copyright Act, 1957 can be changed to cover training-data licensing and the question of the authorship of AI-assisted works.

Keywords: Generative Artificial Intelligence; Copyright; Text and Data Mining; Authorship; Fair Use

1 INTRODUCTION

The field of generative artificial intelligence (AI) has transitioned from the realm of research to infrastructure over the last few years. Prose, software, images, music, you name it, is now written on command by large language models and diffusion models, because they were first trained on very large amounts of existing human-generated text – much of which was copyrighted, and almost none of which explicitly licensed the use of the model's output at the time the original model was created.

This paper focuses on the two copyright issues that generative technology has opened up. The first is an “input question”: Is copying a copyrighted work into a training dataset, and using that to adjust the model's parameters, a violation of the reproduction right, or is it otherwise excused, for example by fair use/fair dealing, etc.? The second is an output question: When using a generative system to create an image, a poem, or code with little or no human creative involvement other than as a prompt, can anything at all be copyrighted, much less by the generator?

* Assistant Professor at Department of Law, Parul Institute of Law, Parul University, Vadodara, Gujarat, India.

** Assistant Professor at Department of Law, Parul Institute of Law, Parul University, Vadodara, Gujarat, India.

Both questions have been the subject of ongoing litigation and regulatory activity in key jurisdictions, and responses are heading in opposite directions with important implications for Indian law and Indian creative industries. This paper examines the state of the law in the USA, where the first substantive judicial decisions on both topics have been issued; in the European Union, where a different regulatory model (opt-out) exists within the framework of the EU AI Act; and in India, where the issue continues to be doctrinally unresolved and where a government policy proposal is pending, as is litigation before the Delhi High Court. The paper asserts that the sticking to the undefined fair-dealing principle by India is a “shaky long-term strategy” and that the better strategy would be to adopt a “calibrated statutory licensing model.”

2 RESEARCH PROBLEM AND RESEARCH METHODOLOGY

The problem of research that is tackled here is two-fold. First, the current copyright doctrine was developed with a different model of authorship and discrete copying in mind; it is not easily adaptable to technology that processes millions of works at once and outputs seemingly independent but nonetheless similar works. In the case of India, there is no text-and-data-mining exception in the copyright law that exists in the European Union and no final judicial ruling as to the scope of Section 52 fair dealing to AI training. Secondly, the Copyright Act, 1957, which governs copyright protection for computer-generated works, provides a definition of an ‘author’ to be the ‘person who causes the work to be created’. However, this fails to distinguish whether a certain amount of human involvement is necessary or not to constitute the work as being made by a ‘person’, especially where a generative system performs most of the expressive act.

This is a comparative and doctrinal paper. It explores key legal texts, such as the Copyright Act, 1957; the United States Copyright Act; Regulation (EU) 2024/1689 (the EU AI Act); and Directive (EU) 2019/790, and evaluates recent case law from the United States and existing Indian case law and policy developments in light of this comparative framework. It does not conduct an empirical analysis of licensing markets or of the designs of licensing architectures; instead, its findings are of a legal-doctrinal nature.

3 THE INPUT PROBLEM – TRAINING DATA AND THE RIGHT OF REPRODUCTION

3.1 Looking for a Place to Start?

Typically, training a generative model involves copying the training corpus, at least temporarily, and putting it in a model’s format, and this requires implication of the reproduction right under most copyright laws, such as Section 14 of the Copyright Act, 1957. The most prevalent defence advanced by AI developers is that this copying is purposeful; that the AI model is neither creating copies for storage nor an attempt to retrieve the original work, but rather is learning some statistical patterns of the original, and that the resulting system is put to a use, such as generating new content, that is not the same as the use of the original as, for example, a news article or photograph.

3.2 The United States: A Doctrine Starting to Crack

In *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, the District of Delaware granted a renewed motion for summary judgment, finding that Ross Intelligence’s use of Thomson Reuters’

copyrighted headnotes to train a rival legal-research tool was not a fair use.¹ Significantly, the court noted that Ross's use was commercial, that it did not involve any meaningful alteration of the material copied, and the significance of market substitution, meaning that such a use was designed to take advantage of the derivative market, for the purpose of teaching competing databases how to function and that Ross's product was intended to compete directly with Westlaw.² While commentators have been careful to point out that Ross's tool was not itself generative and may not be broadly applicable to large language models generating novel content as opposed to excerpts, the ruling is the first judicial setback for the transformative-use argument in the AI-training landscape and is likely to be referred to frequently with respect to the pending generative-AI cases.³

The pending case of greatest interest is perhaps *The New York Times Co. v. Microsoft Corp.* in which the Times claims that millions of its articles were used without permission by OpenAI and Microsoft to train GPT models that generate content that rivals its own journalism.⁴ In April 2025, the district court denied the defendants' motion to dismiss on most of its terms, leaving the Times' central claims to proceed to discovery.⁵ In turn, the case was joined with parallel publisher suits in multidistrict litigation and, at this time, no trial date has been set; the resolution of this case is expected to set the tone for the entire generative-AI industry's exposure to claims involving training data.⁶

There is another parallel (sort of) line of suits involving models that produce images. The plaintiffs in *Andersen v. Stability AI* allege that Stability AI, Midjourney and DeviantArt used copyrighted images to train image models without permission, and Getty Images is suing Stability AI in Delaware for the use of millions of its licensed images, associated captions and metadata, to create Stable Diffusion.⁷ Both cases are still in discovery – and the fact that only the thinnest of margins was found in Getty's favor in the parallel UK case brought by Getty against Stability AI, highlights how unclear the doctrine is even within one litigant's parallel litigation proceedings across jurisdictions.⁸

3.3 The European Union: Opt-Out Approach to Regulation

The European Union has attempted to deal with the same issue in the legislature rather than on an individual basis. The concept of a text and data mining exception, which already exists in Article 4 of Directive (EU) 2019/790, applies only where rightsholders have not expressly reserved their works in an appropriate manner, such as by machine-readable means.⁹ The EU AI Act (Regulation (EU) 2024/1689) adds that providers of general-purpose AI models must implement a policy, thus complying with the EU copyright law, to respect any opt-out reservations and provide publicly a sufficiently detailed summary of

¹ Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., No. 1:20-cv-613-SB, 2025 WL 458520 (D. Del. Feb. 11, 2025).

² Id.; see also Court Rejects Fair Use Defense in AI Copyright Case, Goodwin (Feb. 2025), <https://www.goodwinlaw.com/en/insights/publications/2025/02/alerts-practices-ip-lit-court-rejects-fair-use-defense-in-ai-copyright-case>.

³ Training on Thin Ice: Thomson Reuters v. ROSS and the Future of Fair Use for AI Systems, Ctr. for Art L. (Oct. 2025), <https://itsartlaw.org/art-law/training-on-thin-ice-thomson-reuters-v-ross-and-the-future-of-fair-use-for-ai-systems/>.

⁴ N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023), consolidated as In re: OpenAI, Inc., Copyright Infringement Litig., No. 1:25-md-03143 (S.D.N.Y.).

⁵ NYT v. OpenAI Lawsuit Status 2026, AI Vortex (May 19, 2026), <https://www.aivortex.io/legal/ai-case-law/nyt-v-openai/>.

⁶ AI Copyright Training Data Lawsuits 2026: Status, Timeline, Risk, AI Vortex (May 19, 2026), <https://www.aivortex.io/legal/guides/ai-copyright-training-data-2026-landscape/>.

⁷ Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal.); Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135 (D. Del.).

⁸ The Copyright Battle Over AI-Generated Images: Where Things Stand in 2026, Analytics Insight (2026), <https://www.analyticsinsight.net/news/the-copyright-battle-over-ai-generated-images-where-things-stand-in-2026>.

⁹ Directive 2019/790, art. 4, 2019 O.J. (L 130) 92 (EU).

content used to train the model.¹⁰ It is a very different structure than the “after the fact” fair use litigation in the U.S.: it assumes training is legitimate but allows rightsholders to exercise an affirmative right to withdraw their works from the use of training provided that it is accompanied by a transparency document so that the rightsholders can verify it.

3.4 India: An Unresolved Question

There is no specific text-and-data-mining exception in Indian copyright law. The Copyright Act, 1957 provides an exhaustive list of the various instances of fair-dealing in Section 52, and it has yet to be categorically determined whether AI training of a commercial model would fall under any of them, such as under the protection of private use, research, or incidental use.¹¹ That is exactly the Delhi High Court’s question in *ANI Media (P) Ltd. v. OpenAI, Inc.*, and the news agency ANI has declared that it claims that OpenAI accessed its copyrighted news content without a licence, to train ChatGPT, and that OpenAI itself, similar to its US counterparts, has registered that no harm occurs in India, since the servers on which the training took place are located abroad. The final judgment will be the first definitive judicial pronouncement on the applicability of Indian fair dealing to AI training; and, because of the cross-border reach of Indian copyright law contested by OpenAI, will give important guidance on the cross-border reach of Indian copyright law in the AI context.¹²

While the courts are still deliberating, the executive has taken a separate policy route. In December 2025 the Department for Promotion of Industry and Internal Trade published a working paper, which has proposed a ‘hybrid’ licensing model, where the AI developers may be granted a “licence of sorts” to use copyright content to train AI and a centralized royalty collection body (provisionally named the Copyright Royalties Collective for AI Training) to pay royalties to rightsholders.¹³ Notably, the working paper’s opt-in ‘blanket-licence’ approach (as opposed to the opt-out right in the EU’s text-and-data-mining exception provision) for rightsholders was not widely accepted: although industry bodies separately dissented from certain parts of the proposal on cost basis, creator groups criticised the working paper’s opt-in approach in its current form.¹⁴ The working paper, at this time, is merely a policy proposal because as yet, no Bill to amend the Copyright Act, 1957 has been tabled on the Parliamentary floor and the Ministry of Commerce and Industry has also set up an eight-member expert committee to review the adequacy of the Act in view of generative AI technology.¹⁵

¹⁰ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence, art. 53(1)(c)-(d), 2024 O.J. (L, 1689).

¹¹ Copyright Act, 1957, No. 14, Acts of Parliament, 1957, § 52 (India).

¹² *ANI Media (P) Ltd. v. OpenAI, Inc.*, CS(COMM) 1028/2024 (Del. H.C.) (India); see *The Future of Indian Copyright Legislation in the Wake of ANI Media v OpenAI*, *World Trademark Rev.* (2026), <https://www.worldtrademarkreview.com/guide/india-managing-the-ip-lifecycle/2026/article/the-future-of-indian-copyright-legislation-in-the-wake-of-ani-media-v-openai>.

¹³ Dep’t for Promotion of Indus. & Internal Trade, Working Paper on Artificial Intelligence and Copyright (Part I) (Dec. 2025); AI Training & Copyright in India: DPIIT Hybrid Model Explained, Intepat (May 13, 2026), <https://www.intepat.com/blog/ai-training-copyright-india-dpiit-hybrid-model>.

¹⁴ *The Flaws In India’s Approach To Copyright Reform For AI*, *Live Law* (Jan. 31, 2026), <https://www.livelaw.in/articles/copyright-reform-ai-521269>.

¹⁵ *The future of Indian copyright legislation in the wake of ANI Media v OpenAI*, *supra* note 12.

4 AI-GENERATED WORKS: THE OUTPUT PROBLEM

A more clear judicial response has been given (at least in the United States) to the second question: Can the “creations” of a generative system be the subject of copyright protection, and if so, whose?¹⁶ In March 2026, the Supreme Court declined to grant certiorari in *Thaler v. Perlmutter*, and the human-authorship requirement is now firmly established in the D.C. Circuit at the appellate level — at least on the specific facts of work claimed to have no human creative contribution at all.¹⁷

The ruling doesn’t answer the more difficult and frequent question: is there sufficient human intervention – in prompting, selection, editing – in the creation of an AI’s answer to justify a claim that, at least for those parts, humans acted as the author? The United States Copyright Office’s own guidance has suggested a distinction between merely prompting a system, which is not sufficiently creative to establish a claim to authorship, and material that relies on a human author’s more significant selection, arrangement, or modification of the AI-generated material, in which case, authorship for those parts of the work may be asserted.¹⁸ That line-drawing effort—whether assistive use or just prompting—will likely supplant registration practice and litigation for the foreseeable future, and a related case, *Allen v. Perlmutter*, is likely to test whether the content is assistive use or just prompting in a context more similar to typical prompt-based generation than to the fully autonomous creation claimed by Dr. Thaler.¹⁹

This issue has yet to be put to judicial test in India. The Copyright Act, 1957 already provides Section 2(d)(vi) for computer-generated works, which is peculiar to jurisdictions including India as it defines the author of a literary, dramatic, musical or artistic work made by computer as ‘the person who causes the work to be created’.²⁰ This provision was written with computer-assisted—not computer-autonomous—creation in mind, however, and the words of the statute do not imply copyright functionality for the causing person’s contribution, just causative. No reported Indian decision has tested whether the necessary attributes to constitute a generative-AI prompt satisfy this standard, or even whether the provision has the capacity to support the complete autonomy of large-language-model output at all, and the paper contends that this is equally one of the issues that needs to be addressed either by judicial determination or legislative amendment which provides a minimum standard of human creative participation.

5 EVALUATING REGULATORY MODELS: LITIGATION, EXCEPTION, OR LICENCE

The jurisdictions surveyed are now known to have three general regulatory models. The first, illustrated in the United States, allows decisions to be made case by case in the courts, resulting in fact-sensitive, slow, and – as the contrast between the *Ross Intelligence* decision and the still-pending *New York Times* and *Getty Images* accounts demonstrates – uncertain convergence even in the same legal system. The second type, as with the European Union, sets up a default permission with a far lower hurdle to conform (called opt-out) but also a lesser burden for rightsholders to opt out and a transparency requirement on the AI developer. The third one (proposed, but not yet in effect in India) makes an ‘opt-out’ and individual litigations redundant and provides a compulsory blanket licence based on a centralised collecting

¹⁶ *Thaler v. Perlmutter*, 130 F.4th 1039 (D.C. Cir. 2025).

¹⁷ *Thaler v. Perlmutter*, 130 F.4th 1039 (D.C. Cir. 2025), cert. denied, No. 25-449 (U.S. Mar. 2, 2026).

¹⁸ U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 2: Copyrightability* (Jan. 2025).

¹⁹ *Thaler v. Perlmutter: Human Authors at the Center of Copyright?*, Kluwer Copyright Blog (Apr. 8, 2025), <https://legal-blogs.wolterskluwer.com/copyright-blog/thaler-v-perlmutter-human-authors-at-the-center-of-copyright/>.

²⁰ Copyright Act, 1957, No. 14, Acts of Parliament, 1957, § 2(d)(vi) (India).

society as is provided in the statutory licensing schemes already in existence under Section 31D of the Copyright Act, 1957 for broadcasting.²¹

Every model has varying risk distributions. Litigation-based fair use is unstable for those looking to develop and slow for those whose work is used without their permission to receive payment, but retains the right of the individual rightsholder to outright refuse a licence in the proper case. The opt-out model recently adopted by the European Union helps to limit litigation risk for a developer who has applied the opt-out reservation and transparency requirements, but it may also shift responsibility on to the individual rightsholder that may not have the capacity or awareness to activate the opt-out – the majority of whom are outside of the large publishing houses. This paper considers the blanket-licence model without an opt-out—which is proposed in India to be the system of choice for the country—to be the system’s main weakness as it prioritises AI development without even considering a rightsholder-level opt-out measure.

6 RECOMMENDATIONS

The DPIIT’s hybrid licensing proposal must first be improved before it’s passed by having a rightsholder opt-out clause, similar to the EU’s Digital Single Market Directive, to make the statutory licence, instead, a default rule. But without an opt-out provision, a blanket licence will be vulnerable to constitutional scrutiny as an unreasonable interference with the right to property in copyright (under Article 300A of the Constitution) and it would also prove irreconcilable with the obligations of India in the Berne Convention’s three-step test requiring exceptions to be confined to “certain special cases” which do not “conflict with a normal exploitation of the work.”²²

Second, Section 2(d)(vi) of the Copyright Act, 1957 could be amended by adding a clear definition of what is a ‘computer-generated work’ and by clearly specifying that it can only be a ‘computer-generated work’ where the human contribution is more than a ‘natural-language prompt’ in line with the ‘assistive use’ and ‘mere-prompting’ distinction evolved by the United States Copyright Office, to prevent courts from formulating a definition of the phrase from first principles when the question will certainly arise before them.

Third, require the proposed Copyright Royalties Collective for AI Training to conduct transparency reports in the form of Article 53 of the EU AI Act to ensure that Indian creators can at least have visibility at the ‘dataset’ level if their works have been used by the Collective or not and/or how they have been used, rather than just the ‘hot-potato’ of the royalty payments where the calculation methodology is not transparent and disclosed in the current working paper.

Fourth, as the ANI Media litigation may yet decide the fair-dealing issue before any legislative change is made, any pending decision by the Delhi High Court on this issue should be closely watched as it addresses the territoriality argument – if truly, there is a decision that training of AI models on foreign servers is not subject to Indian copyright jurisdiction, then this ruling, if adopted, would virtually render any copyright regime on such training ineffectual for AI developers with the ability to organise their training on foreign platforms.

²¹ Copyright Act, 1957, No. 14, Acts of Parliament, 1957, § 31D (India).

²² Berne Convention for the Protection of Literary and Artistic Works art. 9(2), Sept. 9, 1886, as revised at Paris, July 24, 1971.

7 CONCLUSION

Generative AI has brought attention to two long maligned copyright holes—the first one pertaining to the limits of what is called a “fair use” to create something new and healthier—a new kind of tool—and the second one has to do with the authorship of whatever that tool creates. Those questions are being answered in court in the United States, and there has been an early decision — Ross Intelligence’s fair-use defence was rejected, and Thaler’s human-authorship requirement was upheld — that the developers of AI might not have hoped would be the case. The EU has opted for a regulatory, opt-out type of response, which gives way for some administrative burden in exchange for regulatory certainty. In India there is a pending case in the Delhi High Court²³ in which only an interim, prima facie ruling has been given so far, a working paper that impacts works generated by a computer machine, yet does not include an opt-out clause for a supporting work, and a 1957 statute, amended in 1994 with computers but not LLMs in mind. The authors of this paper have argued that India should not wait for the courts to weigh in on how to interpret and apply this technology; it cannot be simply worked around with the current interpretation of Section 52 of the Copyright Act, 1957; nor should it merely adopt DPIIT’s proposal for blanket licences.

* * *

²³ ANI Media (P) Ltd. v. OpenAI OPCO LLC, 2026 SCC OnLine Del 5291 (Del. H.C. July 24, 2026) (India) (refusing an interim injunction and holding, prima facie, that the Court had jurisdiction although training took place on foreign servers and that storing ANI’s works to train ChatGPT fell within Section 52(1)(a)(i); the observations are expressly without prejudice to the final adjudication of the suit).