

COPYRIGHT LIABILITY IN GENERATIVE ARTIFICIAL INTELLIGENCE UNDER THE INDIAN LAW

by

Kulvinder

Research Scholar, Department of Law,
Kurukshetra University, Kurukshetra

Abstract

While generative artificial intelligence has made production of text, images, music or other software code and audiovisual works more efficient than ever before, it has also disrupted the foundational structure of copyright law. Such concerns lead to three interrelated questions in Indian copyright doctrine, as follows: does the use of protected works for training purposes infringe rights, when (if ever) do AI-generated outputs constitute infringing copies or adaptations, and can the mere imitation of a human creator's style be legally determined without enclosing ideas, methods and aesthetics? Despite containing provisions on computer-generated works, exclusive rights and infringement, fair dealing and moral rights, The Copyright Act 1957¹ was not drafted to address large scale model training or probabilistic content generation. This paper is doctrinal in nature, inspecting Indian statutes, constitutional provisions as well as case law and policy pertaining to international obligations under copyright laws and comparative jurisdictions including the European Union, Japan and the US. It contends that Indian law ought to reject both extremes: otherwise-free licence for training on protected data, and an over-broad copyright veto over all machine learning. Instead, India should at least move towards a careful calibrated framework that divides copying that takes place during training, substitutive output stage as well as author-specific creative imitation. This paper suggests a new exception for text-and-data-mining by establishing lawful access and transparency/opt-out safeguards through compulsory or collective licensing, plus a human creative control test on AI-assisted authorship, followed by a fault-sensitive liability model for developers, deployers and users.

Keywords: Copyright infringement, Generative AI, Training data, Fair dealing, Indian copyright law, Imitation of creativity

¹Copyright Act 1957 (India) ss 2(d)(vi), 13, 14, 17, 51–55B.

Introduction

Generative AI shifted the focus of copyright law from a familiar copycat question to a more complicated one about learning. Previous digital copyright controversies revolved around issues such as file-sharing, streaming, hyperlinking, database extraction or intermediary liability. Generative AI takes this one step further: not only may datasets copy copyrighted works, they are converting those copies into tokens or embeddings, training models with them and leveraging those models to generate future outputs that may or may not resemble the original work. The legal challenge is to show that the technological process differs from ordinary reading by a human author, and yet not every act of machine learning equates with piracy displacing markets. This issue is now directly relevant to India. More Indian writers, publishers, photographers, musicians and software developers draw on generative AI systems — Indian copyright law has no explicit text-and-data-mining regime. Litigation with regard to the claimed training utilisation of news and publishing content for generative AI framework is forthcoming for final hearing before the Delhi High Court, including ANI Media Pvt Ltd v OpenAI Inc² and similar publishing industry interventions. While these proceedings are significant, they have yet to result in an enduring Indian rule on the merits.

The significance of ANI Media Pvt Ltd v OpenAI Inc (CS(COMM) 1028/2024) extends well beyond its immediate parties. Filed before the Delhi High Court in November 2024, the suit alleges that OpenAI scraped and utilised ANI's copyrighted news content, including material accessible only to subscribers, to train ChatGPT without authorisation or remuneration.³ OpenAI contested the territorial jurisdiction of the Indian court on the basis that its servers are located in the United States, but two court-appointed amici curiae advised that jurisdiction is properly established by reference to the plaintiff's principal place of business in New Delhi.⁴ The Federation of Indian Publishers and the Digital News Publishers Association have both intervened in support of ANI, reflecting broad industry concern that unregulated AI training threatens the economic viability of Indian publishing and journalism.⁵ At the policy level, the Department for Promotion of Industry and Internal Trade (DPIIT) constituted a Committee in

²The Leaflet, 'ANI v OpenAI in the Delhi HC: Everything so far and all that is at stake' (The Leaflet, 1 April 2025) <<https://theleaflet.in/digital-rights/ani-v-openai-in-the-delhi-hc-everything-so-far-and-all-that-is-at-stake>>.

⁴Business Standard, 'Delhi HC has jurisdiction in OpenAI copyright case: Court-appointed experts' (Business Standard, 30 January 2025) <https://www.business-standard.com/companies/news/delhi-hc-copyright-infringement-lawsuit-openai-ani-jurisdiction-ai-training-125013000565_1.html>.

⁵Law Beat, 'Delhi High Court Hears ANI vs OpenAI Copyright Case: DNPA Warns ChatGPT Threatens Survival of News Industry' (Law Beat, 21 August 2025) <<https://lawbeat.in/news-updates/delhi-hc-hears-ani-vs-openai-copyright-case-dnpa-warns-chatgpt-threatens-survival-news-industry-1515625>>.

2025 and released a landmark Working Paper on Generative AI and Copyright (Part I) in December 2025, proposing a ‘One Nation, One License, One Payment’ hybrid licensing model that rejects both a zero-price text-and-data-mining exception and traditional voluntary licensing in favour of a structured framework combining statutory licensing with opt-out and transparency obligations.⁶ The convergence of live litigation and active legislative deliberation confirms that the doctrinal moment this paper addresses is one of immediate practical consequence.

Copyright Act 1957⁷ Section 14 outlines the exclusive rights of copyright owners, including a right of reproduction, communication to the public as well as adaptation and translation. Infringement is set out in Section 51 by reference to an improper use of exclusive rights. Section 52 outlines exceptions such as fair dealing for private or personal use, research, criticism and reporting of current events. Section 57 confers moral rights including the right to attribution. Section 2(d)(vi) provides that in regard to a computer-generated literary, dramatic, musical or artistic work the author is “the person who causes the work to be created.” While these provisions give a baseline, they do not resolve all the questions generative AI raises. Indian copyright doctrine is asked to manage three analytically distinctive stages through provisions drafted for earlier technologies: copying during training, statistical pattern-based outputs, and style replication.

Literature Review

Work on AI and copyright has been structured along three wide themes: authorship, training data and liability for outputs. According to Gervais (2020)⁸, recognising machines as authors is incompatible with the purpose of copyright in incentivising human creation. Ginsburg and Budiardjo (2019)⁹ similarly argue that authorship is still fundamentally linked to human creative choices.

⁶Department for Promotion of Industry and Internal Trade, 'Working Paper on Generative Artificial Intelligence and Copyright (Part I)' (DPIIT, 8 December 2025) Working Paper No P-24029/34/2025-IPR-VII.

⁸DJ Gervais, 'The Machine as Author' (2020) 105 Iowa Law Review 2053.

⁹JC Ginsburg and LA Budiardjo, 'Authors and Machines' (2019) 34(2) Berkeley Technology Law Journal 343.

Indian scholarship has focused especially on Section 2(d)(vi) of the Copyright Act 1957.¹⁰ Chakraborty (2019)¹¹ argues the provision does not address the authorship problem in isolation, as it is ambiguous who “causes” a work to be generated when more than one actor is involved. Samuelson (2023)¹² discusses the large obstacle generative AI presents to copyright, as models can be trained on protected works at scale and derivative outputs lie anywhere between being transformative and substitutive. Lee, Cooper and Grimmelmann (2023)¹³ propose the concept of a generative AI “supply chain” where various copyright implications arise at each step: data scraping, model training, fine-tuning, prompting and output distribution.

This is where Indian fair use scholarship comes in. The Delhi University photocopy litigation — *The Chancellor, Masters and Scholars of the University of Oxford v Rameshwari Photocopy Services*¹⁴ — showed that Indian Courts can read section 52 with a lens around access to knowledge and educational purpose. But that was a case about course packs and education, not for-profit commercial training in AI.

The US Copyright Office’s Part 2 Report of 2025 reaffirmed that copyright requires human authorship, and that fully autonomous AI outputs are not copyrightable regardless of the sophistication of the system or the prompts employed.¹⁵ Part 3 concluded that while computational analysis of works can in principle qualify as transformative, the weight to be accorded to the fourth fair-use factor (market harm) is substantial where AI-generated outputs compete in the same market as the works used for training.¹⁶ The EU AI Act (Regulation (EU) 2024/1689)¹⁷, whose obligations for providers of general-purpose AI models became applicable from 2 August 2025, imposes a copyright compliance policy requirement and a public summary of training content under Article 53(1)(c) and (d), operationalising the text-

¹¹A Chakraborty, 'Authorship of AI Generated Works Under the Copyright Act, 1957' (2019) SSRN <<https://doi.org/10.2139/ssrn.3443006>>.

¹²P Samuelson, 'Generative AI Meets Copyright' (2023) 381(6654) *Science* 158.

¹³Katherine Lee, A Feder Cooper and James Grimmelmann, 'Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain' (2023) arXiv:2309.08133 <<https://arxiv.org/abs/2309.08133>>.

¹⁴*The Chancellor, Masters and Scholars of the University of Oxford v Rameshwari Photocopy Services* 2016 SCC OnLine Del 6229.

¹⁵US Copyright Office, *Copyright and Artificial Intelligence Part 2: Copyrightability* (US Copyright Office, January 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>>.

¹⁶US Copyright Office, *Copyright and Artificial Intelligence Part 3: Generative AI Training, Pre-publication Version* (US Copyright Office, May 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>>.

¹⁷Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (EU AI Act) [2024] OJ L2024/1689, arts 53(1)(c) and (d).

and-data-mining opt-out mechanism established in Article 4(3) of Directive (EU) 2019/790.¹⁸ Japan's Agency for Cultural Affairs issued its General Understanding on AI and Copyright in March 2024, clarifying that Article 30-4 of the Japanese Copyright Act¹⁹ does not apply where the purpose of use includes the 'enjoyment' of the expressive content of a work.²⁰

Research Gap and Original Contribution

A doctrinal framework to navigate generative AI copyright liability is the central deficit in Indian legal scholarship. The broad principle in *R G Anand v Delux Films*²¹ is still good law: similarity of ideas, situations or themes will not be enough unless there is significant copying of expression. Generative AI challenges that principle because a generative model can imitate "the feel" of an author or artist without replicating any individual work. The core contribution of this paper lies in three India-specific layers of liability: training liability, output liability and imitation liability.

Research Questions

- Does the use of copyrighted works as data for training AI violate Indian copyright law, and can section 52 of the Copyright Act 1957²² accommodate such an act?
- When should the output of an AI system be considered an infringing reproduction, adaptation or substantial copy?
- What should Indian law do regarding creative imitation, in particular author-specific style imitation, without protecting abstract style itself?
- Which reforms would build an Indian balance that is in line with constitutional values²³ and international copyright commitments?^{24,25}

Research Objectives

- To analyse the Indian copyright framework relevant to AI training and outputs;

¹⁸Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market [2019] OJ L130/92.

¹⁹Japan Copyright Act (Law No 48 of 1970) art 30-4.

²⁰Agency for Cultural Affairs (Japan), 'General Understanding on AI and Copyright in Japan: Overview' (Agency for Cultural Affairs, March 2024) <https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf>.

²¹*RG Anand v Delux Films* (1978) 4 SCC 118.

²³Constitution of India 1950, arts 19(1)(a), 19(1)(g), 21.

²⁴Berne Convention for the Protection of Literary and Artistic Works (adopted 9 September 1886, as revised) art 9(2).

²⁵Agreement on Trade-Related Aspects of Intellectual Property Rights (adopted 15 April 1994) 1869 UNTS 299 (TRIPS Agreement) art 13.

- To assess relevant Indian case law on concepts like originality, substantial similarity and fair dealing, as well as moral rights;
- To draw contrasts with selected foreign approaches;
- To highlight weaknesses in the current legislation; and
- To suggest pragmatic reforms which India could adopt.

Methodology

This paper uses doctrinal legal research methodology. The research has been carried out on primary legal documents such as the Copyright Act 1957,²⁶ the Constitution of India,²⁷ case law, international treaties and foreign legislation. It also reviews secondary sources including journal articles, government reports, policy papers and legal commentary. Comparative analysis is used selectively. The European Union, Japan and the United States are examined as informative illustrations of three general regulatory alternatives: transparency and opt-out in the EU; broad non-enjoyment use in Japan; and flexible fair use and human authorship doctrine in the US.

Legal Framework

Copyright Act 1957

The primary statute is the Copyright Act 1957.²⁸ Copyright in original literary, dramatic, musical and artistic works, cinematograph films and sound recordings is recognised by Section 13. Section 14 lays down the exclusive rights of the copyright owner: reproduction, issuing copies, communication to the public, adaptation and translation. Even where model training is carried out on temporary or intermediate copies of works, the reproduction right may be involved. Section 51 states that copyright is infringed by a person doing, without licence, any act which the owner of the copyright has exclusive rights to do.

Section 52(1)(a) provides for fair dealing for research or private study, criticism or review, and reporting of current events. Sections 52(1)(b) and 52(1)(c) limit the exemptions from electronic storage to fleeting circumstances. Large-scale commercial AI training does not fit neatly within these provisions. Section 2(d)(vi) provides that the author of a computer-generated work is “the

person who causes such output to be generated.” Originality post-Eastern Book Company v D B Modak²⁹ requires more than labour alone.

The absence of a text-and-data-mining exception from Section 52 is particularly acute given the economic stakes. India’s generative AI market was valued at approximately USD 1.30 billion in 2024 and is projected to grow at a compound annual growth rate of over 15 per cent to reach USD 5.40 billion by 2033.³⁰ At the same time, the Indian Copyright Office issued more than 38,000 registration certificates in 2024 alone.³¹ The three-step test under Article 9(2) of the Berne Convention³² and Article 13 of TRIPS³³ constrains any expansive judicial construction of Section 52: a de facto general exception permitting commercial AI training without opt-out or remuneration mechanisms would struggle to satisfy the requirement that an exception not conflict with the normal exploitation of the work and not unreasonably prejudice the legitimate interests of the author. The Delhi High Court’s purposive approach in *The Chancellor, Masters and Scholars of the University of Oxford v Rameshwari Photocopy Services*³⁴ illustrates that Indian courts are willing to apply Section 52 in context-sensitive and policy-conscious ways.³⁵

Constitutional Framework

The Indian copyright law works within a constitutional order.³⁶ Article 19(1)(a) guarantees freedom of speech and expression. Article 19(1)(g) guarantees the right to practise any profession, or to carry on any occupation, trade or business. Copyright is a mechanism to incentivise expression, but extending that control excessively limits downstream speech, particularly in research, education and innovation. Copyright is a statutory right, not an

²⁹Eastern Book Company v D B Modak (2008) 1 SCC 1.

³⁰IMARC Group, 'India Generative AI Market: Industry Trends, Share, Size, Growth, Opportunity and Forecast 2025–2033' (IMARC Group, 2024) <<https://www.imarcgroup.com/india-generative-ai-market>>.

³¹R Jain and V Yadav, 'Why India Needs a Concrete AI Policy Framework' (The Regulatory Review, 20 February 2026) <<https://www.theregreview.org/2026/02/20/jain-yadav-why-india-needs-a-concrete-ai-policy-framework/>>.

³⁵SpicyIP, 'The Chancellor, Masters and Scholars of the University of Oxford and Ors v Rameshwari Photocopy Services and Ors (DU Photocopy Case)' (SpicyIP) <<https://spicyip.com/resources-links/du-photocopy-case/>>.

absolute natural right. The constitutional balance therefore favours a copyright system that encourages authors but also ensures the sharing of knowledge and technological advancement.

International Obligations

India is a signatory to the Berne Convention³⁷ and the TRIPS Agreement.³⁸ Berne safeguards the economic and moral rights of authors, and Article 9(2) allows for exceptions as long as they are confined to a certain special case and do not (i) conflict with normal exploitation of the work and (ii) unreasonably prejudice the legitimate interests of the author. Article 13 of TRIPS contains an identical standard. Any exception crafted for AI training in India must therefore be carefully calibrated. A calibrated exception — with lawful access, transparency, opt-out and remuneration mechanisms — is more defensible under the three-step test.

Case Law

The Supreme Court in *R G Anand v Delux Films*³⁹ concluded that copyright protects expression and not ideas, concepts, themes, plots or stock situations. Unless an output is a virtually verbatim copy of protected expression, an output that closely mimics the author's general style should not infringe.

The standard for originality was of a purely labour character, which the Supreme Court rejected in *Eastern Book Company v D B Modak*⁴⁰ as necessitating at least some creativity. This is important for works created by AI because outputs that originate purely from a machine should receive no protection unless some form of human creative skill, judgement or selection is involved.

The Eastern Book Company standard carries particular significance for the authorship of AI-assisted works because it closes the 'sweat of the brow' route to copyright that AI developers might otherwise invoke. The Supreme Court, drawing on *CCH Canadian Ltd v Law Society of Upper Canada* and the US Supreme Court's ruling in *Feist Publications Inc v Rural Telephone Service Co*, held that originality requires the author's own intellectual creation, embodying the

author's skill and judgment, and exhibiting more than trivial creativity.⁴¹ Applied to generative AI, this confirms that a purely statistical output, however technically sophisticated, does not attract copyright unless a human author has made sufficiently creative choices in selecting, arranging, or refining it. The consistency between the Eastern Book Company standard and the US Copyright Office's 2025 Part 2 Report, which requires human authorship and holds that merely selecting from AI-generated outputs is insufficient, suggests an emerging international consensus around a human creative control requirement.⁴²

The Delhi University photocopy litigation — *The Chancellor, Masters and Scholars of the University of Oxford v Rameshwari Photocopy Services*⁴³ — showed that Indian Courts can read section 52 with a lens around access to knowledge and educational purpose. While the case will not allow commercial AI training, it illustrates that Indian copyright law is not entirely proprietary. Public interest forms part of its hermeneutic structure.

As developed in *Civic Chandran v Ammini Amma*,⁴⁴ the transformative purpose needs to be considered by courts along with criticism and public interest for purposes of fair dealing under Indian law. India, however, has not embraced the open-ended US fair use doctrine. Consequently, arguments that AI training qualifies as “transformative” can find no acceptance in Indian law without statutory sanction.

Critical Analysis and Discussion

Training Data, Copying and the Limits of Fair Dealing

Machine training is often reproduction done at scale. These copies are legally relevant. Such copying could violate section 14 (the reproduction right)⁴⁵ and section 51 (infringement by copying copyrighted works into a dataset without permission). The more difficult question is whether section 52 provides an exception. Commercial model development is not a good fit for an existing category of fair dealing. The type of “research” under section 52(1)(a) can support forms of machine learning which are academic or not-for-profit, but commercially-

⁴¹Law Times Journal, 'Eastern Book Company and Ors v D B Modak and Ors' (Law Times Journal) <<https://lawtimesjournal.in/eastern-book-company-and-ors-vs-d-b-modak-and-ors/>>.

⁴²US Copyright Office (n 5).

⁴⁴*Civic Chandran v Ammini Amma* 1996 PTC (16) 329 (Ker).

focussed foundation-model development does not constitute private research in the statutory sense.

The conclusion should not be that every act of AI training needs a separate licence from each right holder. That would create prohibitive transaction costs, favour incumbent technology firms and make socially beneficial research difficult. A specific text-and-data-mining provision should be included in Parliamentary legislation drawing a line between lawful-access and unlawful-source training.

The December 2025 DPIIT Working Paper reached a broadly consistent conclusion, recommending a hybrid licensing model that combines a statutory right for certain computational analysis uses with binding obligations on commercial developers to maintain dataset documentation, honour opt-out reservations, and participate in sector-specific collective licensing arrangements.⁴⁶ Critics, including Takshashila Institution, have cautioned that mandatory collective licensing risks concentrating licence revenue in large aggregators while providing minimal benefit to individual authors, and that compliance costs could disproportionately burden Indian AI startups relative to well-resourced multinational platforms.⁴⁷ These tensions confirm that India requires a bespoke framework that accounts for the asymmetry between large foreign AI developers and the predominantly individual and small-publisher character of India's domestic creator economy.

Output Liability and Human Responsibility

Output should not be infringing merely because the dataset used to train a model included copyrighted works. There must be a compelling connection between the original work and any allegedly infringing output. Indian courts ought to inquire whether the output incorporates a substantial amount of protected expression. Outputs derived from genre conventions, themes, facts, public domain material or style should not infringe. This methodology is in keeping with R G Anand⁴⁸ and the idea-expression dichotomy.

⁴⁶Department for Promotion of Industry and Internal Trade (n 4).

⁴⁷Takshashila Institution, 'Policy Advisory: Working Paper on Generative AI and Copyright (Part 1)' (Takshashila Institution, 6 February 2026) <<https://takshashila.org.in/content/publications/20260206-generative-ai-and-copyright-consultation.html>>.

The actor-sensitive analysis aligns with the ‘supply-chain’ framework proposed by Lee, Cooper and Grimmelmann,⁴⁹ who argue that copyright liability in generative AI must be assessed at each distinct stage: data scraping, model training, fine-tuning, prompting, and output distribution. This staged analysis is consistent with the structure of Section 51 of the Copyright Act 1957,⁵⁰ and with the holding in *Entertainment Network (India) Ltd v Super Cassettes Industries Ltd*⁵¹ that the intermediary’s knowledge of and control over infringing activity is material to the assessment of secondary liability.

Style, Imitation and the Limits of Copyright

There is no protection for style per se. Copyright is meant to protect expression, not technique, method, genre, mood or aesthetic. The right approach is to preserve the idea-expression divide established in *R G Anand v Delux Films*⁵² and use other means for specific harms. Where AI-generated outputs falsely suggest that a work is by a particular author, moral rights under Section 57,⁵³ passing-off, consumer protection or personality-rights principles may apply.

Authorship and Ownership of AI-Generated Works

Section 2(d)(vi)⁵⁴ names an author for computer-generated works but leaves unanswered the fundamental question of what is original in complex AI systems. The *Eastern Book Company*⁵⁵ originality standard forecloses the argument that considerable labour and capital invested in training a foundation model can, of itself, generate copyright in the model’s outputs. Indian law ought to follow a human creative control test: only human-authored material should be protected, not the raw output of an autonomous agent.

Weaknesses in the Current Indian Framework

⁵¹*Entertainment Network (India) Ltd v Super Cassettes Industries Ltd* (2008) 13 SCC 30.

There are five weaknesses in the existing framework.

First, text-and-data-mining (TDM) is not explicitly exempted in section 52⁵⁶ of the Copyright Act 1957. Courts might be compelled to stretch existing categories, generating uncertainty.

Secondly, section 2(d)(vi) is insufficient for generative AI. It does not define what causes a work to be created, does not address creative control where many actors are involved, or ownership.

Third, there are no transparency requirements in Indian law about the datasets on which AI is trained. If authors are unable to find out whether their works were used, they cannot meaningfully enforce rights.

Fourth, existing remedies are ill-suited to AI systems. It may be difficult or disproportionately excessive to delete traces of works from a trained model, and where unauthorised training creates ongoing commercial value, damages alone may not suffice.

Fifth, the law draws an initial but inadequate distinction between non-commercial research and public-interest AI development on the one hand, and for-profit foundation models and tools designed to infringe on the other.

Comparative Perspective

European Union

The European Union provides a kind of template. Directive (EU) 2019/790⁵⁷ establishes text-and-data-mining exceptions. Article 3 is applicable to research organisations and cultural heritage institutions. Article 4 is drafted more broadly, allowing text and data mining of legally accessible works, while permitting right holders to opt out through machine-readable means. Providers of general-purpose AI models are given transparency duties as well as policies to adhere to copyright law under the EU AI Act (Regulation (EU) 2024/1689).⁵⁸

Under Article 53(1)(c) and (d), GPAI model providers must implement a policy to comply with Union copyright law, including respecting machine-readable opt-out reservations under Article 4(3) of Directive (EU) 2019/790, and must publish a sufficiently detailed summary of training

content using a template issued by the AI Office in July 2025. A GPAI Code of Practice, also published in July 2025, provides operational guidance for compliance.⁵⁹ As Martens has observed, the EU framework remains caught between the demands of copyright law and the competitive pressure to ensure access to frontier AI models.⁶⁰ For India, the EU model's most transferable feature is the combination of a conditional TDM right (tied to lawful access and machine-readable opt-out) with a public training-data disclosure requirement.

Japan

Japan is known as an AI-friendly country when it comes to copyright law because of the non-enjoyment purposes laid down in Article 30-4 of the Japan Copyright Act,⁶¹ which permits use of works for information analysis. This enables general machine learning where the intent is not to enjoy the expressive content of a work.

The March 2024 General Understanding on AI and Copyright, issued by the Agency for Cultural Affairs' Legal System Subcommittee, clarified that the 'non-enjoyment purpose' requirement is assessed in relation to the specific work being used, and that uses involving 'overlearning' of protected expression — where a model is fine-tuned to reproduce the distinctive expressive style of a specific author on demand — fall outside the exception even under Japan's permissive regime.⁶² This significantly narrows the practical distance between the Japanese and EU models: both permit computational analysis of works at the training stage but restrict outputs that function as substitutes for protected expression.

United States

Many of the legal disputes over AI training in the United States are framed through fair use. The US Copyright Office⁶³ vigorously upholds human authorship. Copyright claims arising in works produced without sufficient human creative agency have been rejected by the US

⁵⁹Clifford Chance, 'Copyright Compliance Under the EU AI Act for GPAI Model Providers' (Clifford Chance, November 2025) <<https://www.cliffordchance.com/insights/resources/blogs/ip-insights/2025/10/copyright-compliance-under-the-eu-ai-act-for-gpai-model-providers.html>>.

⁶⁰U Martens, 'The European Union is still caught in an AI copyright bind' (Bruegel, 10 September 2025) <<https://www.bruegel.org/analysis/european-union-still-caught-ai-copyright-bind>>.

⁶²Agency for Cultural Affairs (Japan) (n 8); Hugh Stephens, 'Japan's Text and Data Mining (TDM) Copyright Exception for AI Training: A Needed and Welcome Clarification from the Responsible Agency' (Hugh Stephens Blog, 10 March 2024) <<https://hughstephensblog.net/2024/03/10/japans-text-and-data-mining-tdm-copyright-exception-for-ai-training-a-needed-and-welcome-clarification-from-the-responsible-agency/>>.

Copyright Office and federal courts, as demonstrated in *Thaler v Perlmutter*.⁶⁴ The American model is helpful but not entirely transferable. India has a more specific section 52 than US fair use. Indian courts cannot write off commercial AI training as “transformative” unless the statutory text so allows.

Findings

This paper makes seven findings.

- Reproduction in pursuit of AI training, although constituting copying, should not be considered actionable infringement in every case under the Copyright Act 1957.
- Section 52 does not offer a distinct defence for non-commercial AI training at scale. Its current text is inadequate for protecting training on foundation models, although it may protect some limited research or incidental technical uses.
- The test should be one of substantial similarity and protected expression. The idea-expression distinction, as affirmed in *R G Anand v Delux Films*,⁶⁵ remains central.
- Straightforward style imitation may not be an infringement of copyright. By contrast, market substitution, false attribution, and authorship exploitation may be covered by moral rights under Section 57, passing off, consumer law or targeted AI regulation.
- Section 2(d)(vi) should not be treated as conferring automatic copyright on autonomous AI outputs. Human creative control, per the *Eastern Book Company*⁶⁶ standard, should be mandated.
- The current Indian architecture lacks accountability, documentation and incentive regimes for AI training.
- Comparative law moves towards a middle ground: legal computational analysis should be permitted, but for-profit AI systems must meet obligations of transparency, opt-out, rights protection and, in some cases, compensation.

Recommendations and Reform Proposals

Specific Text-and-Data-Mining Exception

India needs to update section 52 of the Copyright Act 1957 by adding an exception for text and data mining. The exception shall be available only in relation to works lawfully accessed by the user. It must enable reproducing any process necessary to conduct a computational analysis,

⁶⁴*Thaler v Perlmutter* No 23-5233 (US Court of Appeals for the District of Columbia Circuit, 18 March 2025).

eg indexing, tokenisation and feature extraction. It must not encompass works obtained from piratical sources, hacked databases or circumvention of paywalls.

Differentiating between Research and Commercial Training

The law must make non-commercial research, education, libraries and public-interest institutions a wider privileged use. Additional obligations should apply to developers of commercial foundation models, including dataset documentation, opt-out compliance, and licensing mechanisms.

Introduce Transparency Duties

Large generative AI systems released in India must provide meaningful summaries of the categories of data used for training. Internal records should include sufficient supporting documentation to facilitate audit by an appointed authority or court. Transparency must provide enough information for right holders to assess likely use, without requiring disclosure of individual works or trade secrets.

Opt-Out and Rights-Reservation Mechanism

Machine-readable metadata, platform registries or collective management systems should allow right holders to reserve rights against commercial AI training. Opt-out should not trump non-commercial research uses or public-interest exemptions, consistent with the framework established under Article 4(3) of Directive (EU) 2019/790.⁶⁷

Collective Licensing for Valuable Sectors

India should promote collective or extended collective licensing in high-value sectors such as news, books, music and stock images. Licensing needs to be made clear, open and free from bias towards large technology corporations.

Human Creative Control Test

The Copyright Act should be amended to make clear that copyright in AI-assisted works subsists only in human creativity, consistent with the originality standard in *Eastern Book Company v D B Modak*.⁶⁸ Courts should take into account specificity of the prompt, iterative

control, selection, editing, arrangement and expressive judgement. Raw autonomous output must not be afforded full copyright protection.

Fault-Sensitive Liability Regime

Developers who knowingly use unlawful datasets, design systems to reproduce protected expression, or market infringing imitation tools should be liable. Where users create infringing outputs for the purpose of exploiting them, this should also be actionable. Deployers who deploy AI systems commercially with knowledge that their systems infringe or pose a risk of infringement should be held liable, consistent with the principle established in Entertainment Network (India) Ltd v Super Cassettes Industries Ltd.⁶⁹

Protection Against False Attribution and Damage to Reputation

India should enhance protection of claims of AI-generated works being wrongly attributed to human authors. Section 57 of the Copyright Act 1957⁷⁰ may assist in some situations, but further clarification through rules or guidance would be advisable. Consumer protection and passing-off principles should deal with misleading AI outputs presented in commerce as works of specific creators.

Conclusion

This paper explored whether Indian copyright law will provide a remedy for AI training data, generated outputs and creative imitation. Training could entail reproduction, but the existing section 52⁷¹ does not clearly settle commercial foundation-model training. A statutory text-and-data-mining exception is needed. The paper contended that outputs should be evaluated by substantial similarity and fair dealing. Not every output of a model trained on copyrighted material should be infringing. Outputs with significant reproduced protected expression should be subject to liability. Copyright should not cover bare style: the idea-expression divide established in R G Anand v Delux Films⁷² needs to be preserved. Where there is false attribution or targeted market substitution, legal solutions are available through moral rights

under Section 57,⁷³ passing off, consumer protection or positive law regulation. The paper argues for an Indian model based on lawful access, transparency, opt-out, collective licensing, human creative control and fault-sensitive liability. The core contribution is a three-tier doctrinal framework: training liability, output liability and imitation liability.

Bibliography

A Primary Sources

Legislation

Copyright Act 1957 (India) ss 2(d)(vi), 13, 14, 17, 51–55B

Constitution of India 1950, arts 19(1)(a), 19(1)(g), 21

Japan Copyright Act (Law No 48 of 1970) art 30-4

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market [2019] OJ L130/92

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L2024/1689

International Instruments

Berne Convention for the Protection of Literary and Artistic Works (adopted 9 September 1886, as revised)

Agreement on Trade-Related Aspects of Intellectual Property Rights (adopted 15 April 1994) 1869 UNTS 299 (TRIPS Agreement) art 13

Cases

ANI Media Pvt Ltd v OpenAI Inc & Anr CS(COMM) 1028/2024 (Delhi High Court)

Civic Chandran v Ammini Amma 1996 PTC (16) 329 (Ker)

Eastern Book Company v D B Modak (2008) 1 SCC 1

Entertainment Network (India) Ltd v Super Cassettes Industries Ltd (2008) 13 SCC 30

Feist Publications Inc v Rural Telephone Service Co 499 US 340 (1991)

R G Anand v Delux Films (1978) 4 SCC 118

Super Cassettes Industries Ltd v Hamar Television Network Pvt Ltd (2011) SCC OnLine Del 2726

The Chancellor, Masters and Scholars of the University of Oxford v Rameshwari Photocopy Services 2016 SCC OnLine Del 6229

Thaler v Perlmutter No 23-5233 (US Court of Appeals for the District of Columbia Circuit, 18 March 2025)

B Secondary Sources

Books and Journal Articles

A Chakraborty, 'Authorship of AI Generated Works Under the Copyright Act, 1957' (2019) SSRN <<https://doi.org/10.2139/ssrn.3443006>>

TW Dornis and S Stober, 'Copyright Law and Training Generative AI' (2025) arXiv:2502.15858 <<https://arxiv.org/abs/2502.15858>>

DJ Gervais, 'The Machine as Author' (2020) 105 Iowa Law Review 2053

JC Ginsburg and LA Budiardjo, 'Authors and Machines' (2019) 34(2) Berkeley Technology Law Journal 343

K Lee, A Feder Cooper and J Grimmelmann, 'Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain' (2023) arXiv:2309.08133 <<https://arxiv.org/abs/2309.08133>>

P Samuelson, 'Generative AI Meets Copyright' (2023) 381(6654) Science 158

AW Torrance and B Tomlinson, 'Getting the Most from Training: Artificial Intelligence, Copyright and Fair Training' (2023) arXiv:2305.03720 <<https://arxiv.org/abs/2305.03720>>

Reports and Policy Documents

Agency for Cultural Affairs (Japan), 'General Understanding on AI and Copyright in Japan: Overview' (Agency for Cultural Affairs, March 2024) <https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf>

Clifford Chance, 'Copyright Compliance Under the EU AI Act for GPAI Model Providers' (Clifford Chance, November 2025) <<https://www.cliffordchance.com/insights/resources/blogs/ip-insights/2025/10/copyright-compliance-under-the-eu-ai-act-for-gpai-model-providers.html>>

Department for Promotion of Industry and Internal Trade, 'Working Paper on Generative Artificial Intelligence and Copyright (Part I)' (DPIIT, 8 December 2025) Working Paper No P-24029/34/2025-IPR-VII

IMARC Group, 'India Generative AI Market: Industry Trends, Share, Size, Growth, Opportunity and Forecast 2025–2033' (IMARC Group, 2024) <<https://www.imarcgroup.com/india-generative-ai-market>>

R Jain and V Yadav, 'Why India Needs a Concrete AI Policy Framework' (The Regulatory Review, 20 February 2026) <<https://www.theregreview.org/2026/02/20/jain-yadav-why-india-needs-a-concrete-ai-policy-framework/>>

Law Beat, 'Delhi High Court Hears ANI vs OpenAI Copyright Case: DNPA Warns ChatGPT Threatens Survival of News Industry' (Law Beat, 21 August 2025) <<https://lawbeat.in/news-updates/delhi-hc-hears-ani-vs-openai-copyright-case-dnpa-warns-chatgpt-threatens-survival-news-industry-1515625>>

- U Martens, ‘The European Union is still caught in an AI copyright bind’ (Bruegel, 10 September 2025) <<https://www.bruegel.org/analysis/european-union-still-caught-ai-copyright-bind>>
- Parliamentary Standing Committee on Commerce, ‘161st Report: Review of the Intellectual Property Rights Regime in India’ (Rajya Sabha, 2021) <<https://prsindia.org/>>
- SpicyIP, ‘The Chancellor, Masters and Scholars of the University of Oxford and Ors v Rameshwari Photocopy Services and Ors (DU Photocopy Case)’ (SpicyIP) <<https://spicyip.com/resources-links/du-photocopy-case>>
- Hugh Stephens, ‘Japan’s Text and Data Mining (TDM) Copyright Exception for AI Training’ (Hugh Stephens Blog, 10 March 2024) <<https://hughstephensblog.net/2024/03/10/japans-text-and-data-mining-tdm-copyright-exception-for-ai-training-a-needed-and-welcome-clarification-from-the-responsible-agency/>>
- Step toe, ‘EU AI Act Obligations for GPAI Models Now Applicable’ (Step toe, 5 August 2025) <<https://www.step toe.com/en/news-publications/step tech toe-blog/eu-ai-act-obligations-for-gpai-models-now-applicable.html>>
- Takshashila Institution, ‘Policy Advisory: Working Paper on Generative AI and Copyright (Part 1)’ (Takshashila Institution, 6 February 2026) <<https://takshashila.org.in/content/publications/20260206-generative-ai-and-copyright-consultation.html>>
- The Leaflet, ‘ANI v OpenAI in the Delhi HC: Everything so far and all that is at stake’ (The Leaflet, 1 April 2025) <<https://theleaflet.in/digital-rights/ani-v-openai-in-the-delhi-hc-everything-so-far-and-all-that-is-at-stake>>
- US Copyright Office, Copyright and Artificial Intelligence Part 2: Copyrightability (US Copyright Office, January 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>>
- US Copyright Office, Copyright and Artificial Intelligence Part 3: Generative AI Training, Pre-publication Version (US Copyright Office, May 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>>