GENERATIVE AI AND COPYRIGHT LAW IN INDIA: AUTHORSHIP, FAIR DEALING AND THE NEED FOR A TEXT-AND-DATA-MINING EXCEPTION

Nirmit Singh*

ABSTRACT

Generative AI and copyright law in India are increasingly at odds. Data-hungry models depend on large amounts of protected works for training. They also generate outputs that are similar to those created by humans. This paper examines how India's Copyright Act of 1957, which primarily focuses on human authors, struggles to address the automated copying that occurs during training and the issues of authorship and ownership related to AI-assisted and fully AI-generated works. The paper first examines the laws and case rulings. It argues that Section 2(d) does not recognise non-human authorship. A limited, user-as-author approach could better serve our constitutional commitments to creativity and free speech. Next, the paper explores whether large-scale text-and-data mining (TDM) for AI training can be supported under Section 52's fair dealing. It concludes that current exceptions are too limited to cover commercial generative AI systems effectively. By considering comparative views and recent Indian cases, such as ANI v OpenAI, this paper ultimately advocates for a specific TDM exception. This exception should include clear safeguards that (a) maintain incentives for human creators, (b) allow regulated AI training, and (c) clarify the line between AI-assisted creativity and unprotected, fully automatic outputs.

Keywords: Generative AI, Copyright Law, Authorship, Fair Dealing, Text-and-Data Mining.

INTRODUCTION

Generative AI systems, like large language models and image generators, are trained on extensive collections of text, images, audio, and video, much of which is protected by copyright. These systems can produce outputs that look like human-made works, including essays, artwork, music, and software code. Indian copyright law, created for a world focused

*BA LLB, FIRST YEAR, BHARATI VIDYAPEETH, NEW LAW COLLEGE, PUNE.

on humans and drafted before AI, now needs to address both automated copying during model training and the status of content generated by AI.[1]

The Copyright Act of 1957 defines "author" in human terms. It grants exclusive rights, such as reproduction, adaptation, and communication to the public, to right holders. These rights are mainly enforced through Sections 14 and 51. Section 52 includes a flexible list of "fair dealing" and other exceptions. Scholars and policymakers are currently examining these exceptions in the context of data-hungry AI systems and TDM techniques. This article explores three related questions: the authorship of AI outputs, fair dealing and the doctrine of transformative use during training, and the necessity for a specific statutory transformative use exception in India.[2]

GENERATIVE AI AND INDIAN COPYRIGHT BASICS

Indian law currently does not recognise non-human authorship. The Copyright Act assigns authorship only to natural persons in Section 2(d). As a result, purely AI-generated content, where there is no significant human input, falls into a grey area. Many commentators argue that such works are not protected by copyright under Indian law.[3]

AUTHORSHIP AND OWNERSHIP OF AI-GENERATED WORKS

Human-Centric Authorship in Statute and Case Law: The scheme of Section 2(d) clearly states that authors must be human, whether they create literary, artistic, musical, or cinematographic works. Indian courts have strengthened this understanding by highlighting that human skill and judgment are necessary for originality and authorship. In a recent case involving the RAGHAV AI system, the Delhi High Court did not recognise an AI tool as an "author" under Section 2(d). The court held that authorship must be based on human involvement. This aligns with a broader global agreement in many areas that AI systems cannot hold copyright, even if they create content on their own.

A clear distinction is important: one where humans use AI as a tool or assistive technology, and another where AI operates independently to create content without much human control or

[1] Kailash Chauhan, 'Generative AI, Text & Data Mining and the Fair Dealing Doctrine: Examining the New Problem with the Old Regime' (2025) 30(1)
JIPR https://or.niscpr.res.in/index.php/JIPR/article/view/12652 accessed 21 December 2025.
[2] Kailash Chauhan, 'Generative AI, Text & Data Mining and the Fair Dealing Doctrine: Examining the New Problem with the Old Regime' (2025) 30(1)
JIPR https://or.niscpr.res.in/index.php/JIPR/article/view/12652 accessed 21 December 2025.
[3] Maheshwari & Co, 'AI Copyright Law India: Ownership Explained' (Maheshwari and Co, 15 July 2025) https://www.maheshwariandco.com/blog/ai-copyright-law-india/ accessed 21 December 2025.

input. In the first case, courts and analysts argue that the human user who gives instructions, edits the output, and applies creative judgment could be seen as the author, similar to how digital tools or software are regarded. In the second case, the output might not have a human author under the law, which means it could be unprotected or considered part of the public domain.[4]

Ownership and the "User as Author" Approach: Scholars studying Indian law suggest that ownership of works created with AI should mainly belong to the person who provides the intellectual input and directs the creative process. They propose changing Section 2(d) to clarify authorship in situations involving AI. This could involve using language similar to that in the UK, which considers the person who organises the creation of computer-generated works as the author. This change would support the human-centred basis of Indian copyright while making sure that AI-assisted creativity is protected.

At the same time, commentators caution against giving exclusive rights to fully automated AI outputs without significant human input. This could lead to overly broad monopolies that ignore the reason for rewarding human creativity. A clear legal definition that differentiates between AI-assisted and AI-generated works, while tying authorship to clear human contribution, would fit both the structure of the Act and the principles of free speech in the Constitution.[5]

FAIR DEALING AND GENERATIVE AI

Scope of Fair Dealing under Section 52: Section 52 of the Copyright Act states that some actions do not count as infringement. These include fair dealing with a work for reasons like private or personal use, including research, criticism or review, and reporting on current events. Indian courts have created a flexible, factor-based approach to fair dealing. They look at the purpose and character of the use, the amount and significance of what is used, and the impact on the market for the original work. This method is often compared to the four-factor fair use test in the United States, but it is still based on the specific purposes outlined in Section 52.

[4] Maheshwari & Co, 'AI Copyright Law India: Ownership Explained' (Maheshwari and Co, 15 July 2025) https://www.maheshwariandco.com/blog/ai-copyright-law-india/ accessed 21 December 2025.
[5] Kailash Chauhan, 'Generative AI, Text & Data Mining and the Fair Dealing Doctrine: Examining the New Problem with the Old Regime' (2025) 30(1)
JIPR https://or.niscpr.res.in/index.php/JIPR/article/view/12652 accessed 21 December 2025.

Limits of Fair Dealing for AI Training: Analyses of Indian law show that large-scale, automated copying for AI training does not clearly match any of the purposes listed in Section 52. This is especially true when the final product is a commercial generative AI tool. While "private research" might cover small-scale, non-commercial experimentation, using millions of copyrighted works for global AI services seems quite different from the traditional idea of private use. Additionally, fair dealing in India is not a broad defence; it is tied to specific categories and does not operate as a general fairness test like US fair use.

ANI vs. OpenAI is a significant copyright infringement lawsuit. It was filed by the Indian news agency Asian News International (ANI) against OpenAI in India's Delhi High Court. ANI claims that OpenAI's ChatGPT used ANI's news content without permission to train its AI models. This has led to economic damage and misrepresentation. This case is important for Indian law. It will test fair use, copyright, and the future of AI's access to data. The outcome will have major implications for content creators and AI development.[6]

TEXT-AND-DATA-MINING AND THE LEGISLATIVE GAP

TDM and International Approaches: Text-and-data mining involves automated techniques to analyse large amounts of text and data to find patterns, trends, or other insights. It plays a key role in training modern AI systems. Several regions have set up specific TDM exceptions that allow for temporary reproduction and analysis of copyrighted works for machine learning. These exceptions often come with conditions, such as lawful access, non-commercial use, or opt-out options for rights holders. For instance, the EU's Directive on Copyright in the Digital Single Market offers limited TDM exceptions, while Japan has taken a broader, technology-friendly approach that permits TDM for any purpose as long as certain safeguards are in place.

Comparative studies indicate that flexible TDM exceptions can encourage AI innovation, especially in countries with various languages and cultural content. This flexibility still allows authors to maintain control over expressive uses and market alternatives. However, the specific details of these exceptions, including who benefits, whether commercial TDM is allowed, and

[6] Kailash Chauhan, 'Generative AI, Text & Data Mining and the Fair Dealing Doctrine: Examining the New Problem with the Old Regime' (2025) 30(1)
JIPR https://or.niscpr.res.in/index.php/JIPR/article/view/12652 accessed 21 December 2025.

how rights holders can opt out or seek payment, remain points of contention in policy discussions around the world.[7]

THE EMERGING DPIIT AI–COPYRIGHT FRAMEWORK

Proposed Regulatory Architecture: Recent reports indicate that the Department for Promotion of Industry and Internal Trade (DPIIT) is developing a specialised AI copyright framework, rather than utilising a broad TDM exception, as seen in the EU or Japan. Draft proposals reportedly include mandatory access rules, state-set royalty rates, and centralised systems for collecting and distributing payments for the use of copyrighted works in AI training. Unlike global TDM models that offer blanket exceptions with opt-out rights, the proposed Indian approach seems to mix required access to works with regulated payments and state supervision.

Constitutional and Policy Concerns: Legal experts contend that an excessively dirigiste AI-copyright plan might be at odds with both the fundamental framework of the Copyright Act and more general constitutional principles. If mandatory access or licensing is not properly tailored and justified, it could be challenged for disproportionately interfering with the rights of authors and intermediaries. However, a lack of legal clarity regarding AI training could stifle innovation and put Indian developers at a disadvantage in comparison to their international.[8]

CONCLUSION

Generative AI has highlighted long-standing tensions in Indian copyright law. This tension concerns the balance between protecting authors and supporting technological progress, especially regarding authorship, fair dealing, and large-scale text and data mining. The current focus on human authorship does not correctly recognise AI systems as authors. However, it needs clarification to deal with AI-assisted work. We must ensure that human creative contributions get rewarded without giving monopolies to fully automated outputs. Relying on Section 52 fair dealing to justify large-scale AI training is legally questionable and practically

[7] Veritas Legal, 'Text and Data Mining – Decoding Copyright Challenges in India' (Veritas Legal, 19 October 2023) https://www.veritaslegal.in/legal-update-text-and-data-mining-decoding-copyright-challenges-in-india/ accessed 25 December 2025.
[8] ABC Live, 'Explained: The Constitutional Risks in DPIIT's AI Copyright Plan' (ABC Live, 8 December 2025) https://abclive.in/2025/12/09/dpiit-ai-copyright-plan/ accessed 27 December 2025.

unclear. This indicates a need for a specific statutory TDM exception that is clear, fair, and supports innovation.

As DPIIT and other organisations develop India's AI and copyright policy, the challenge will be to avoid both over-regulation that burdens a growing sector and under-regulation that weakens author rights and constitutional values. A well-structured TDM exception, along with clearer rules on AI-assisted authorship and reasonable protections for creators, provides a solid approach to modernising Indian copyright law in light of generative AI while promoting the public interest in knowledge, culture, and technological progress.