AICYBERLEX NEXUS

Journal of Artificial Intelligence, Cyber & Contemporary Law

Volume I · **Issue I** · **Jul–Sep 2026** · Article 5
ISSN: Applied for (National ISSN Centre, NIScPR, India) · e-ISSN: TBA
Peer-reviewed · Open Access · https://aicyberlexnexus.com

ARTICLE 5

Who Is the Author? Generative AI and Copyright in India

Vikram

ABSTRACT

Generative AI raises two separate copyright questions. The first is about output: can a work produced by a machine be protected, and if so, who owns it? The second is about input: is it lawful to train a model on millions of copyrighted works without asking their owners? Indian law has no settled answer to either. The Copyright Act, 1957 contains an unusual provision on computer-generated works, but its wording predates modern AI, and the training question is now before the Delhi High Court. This paper sets out the statutory position, compares it with approaches in the United States, the United Kingdom and the European Union, and considers the main reform options, including a text-and-data-mining exception, a licensing model and a transparency duty. It argues that India should separate the two questions, keep human creativity at the centre of authorship, and adopt a licensing-based answer on training that pays creators without halting research.

KEYWORDS: generative AI; copyright; authorship; computer-generated works; text and data mining; fair dealing

1. Introduction

A novelist types a few prompts and receives a chapter. A designer asks an image tool for a poster. A musician generates a backing track. All three want to know the same thing: do I own this? And on the other side, a news agency, a publisher and a record label want to know something else: did the company that built the tool have the right to learn from their work? These are different questions, and much confusion in public debate comes from mixing them. One is about the end of the pipeline, and the other is about the beginning. Different legal rules apply to each, and the commercial stakes are different. India is an interesting place to ask them. The Copyright Act, 1957 already says something about computer-generated works, which few other countries' statutes do. And the first major Indian case on AI training is pending. The paper therefore tries to take stock before the courts and Parliament move. Part 2 sets out the statutory framework. Part 3 considers the ownership of outputs. Part 4 deals with training data. Part 5 examines reform options, and Part 6 concludes.

2. Authorship under the Copyright Act, 1957

Copyright in India protects original literary, dramatic, musical and artistic works, along with films and sound recordings (section 13). Originality has been explained by the Supreme Court in Eastern Book Company v D.B.

Modak, which rejected the older "sweat of the brow" approach and required a "modicum of creativity" and the exercise of skill and judgment.[1] This approach points to human intellectual effort as the basis of protection.

1

The Act also defines "author". For a literary, dramatic, musical or artistic work that is computer-generated, section 2(d)(vi) names "the person who causes the work to be created" as the author.[2] The provision was added in 1994, long before generative AI. It was meant for outputs such as computer-produced maps and designs, where a person set the process in motion and the machine did the rest. First ownership then follows the author under section 17, subject to the usual exceptions for employment and commissioned work. Copyright in a literary work lasts for the author's life plus sixty years (section 22), which creates an obvious problem when the "author" is defined by a causal act rather than a life, a point discussed below.

3. Who owns what the model produces?

3.1 The Indian position On its face, section 2(d)(vi) seems to help the user who writes a prompt: she is the person who "causes" the work to be created. But the wording raises three difficulties. First, the word "computer-generated" has no definition in the Act, and it is unclear whether a work made by a large generative model, which draws on enormous training data and produces results the user did not foresee, falls within it. Second, the section does not displace the originality requirement. If the output has no human creative contribution, a court may well find no original work to protect. A short generic prompt ("a sunset over mountains") adds little to the picture that comes out. A detailed series of choices, edits and selections is more likely to show the skill and judgment Eastern Book Company demands. Third, the term of protection in section 22 is tied to the life of an author, and fits a machine-generated work badly. There is also the registration episode. In 2020 the Indian Copyright Office registered an artwork listing an AI system as co-author alongside a human, and later issued a notice proposing to withdraw the registration. The episode shows that the Office itself has not settled its position. 3.2 The United States The American approach is stricter. The Copyright Office has said that works need human authorship, and that prompts alone do not usually give a user enough control over the output to count as authorship, although human selection, arrangement and modification of AI material can be protected.[3] In Thaler v Perlmutter the Court of Appeals for the District of Columbia Circuit confirmed that a work created autonomously by an AI system, with no human author, could not be registered.[4] 3.3 The United Kingdom The UK, like India, has a computer-generated works provision. Section 9(3) of the Copyright, Designs and Patents Act 1988 names as author the person by whom "the arrangements necessary for the creation of the work are undertaken", and gives protection for fifty years.[5] The government has consulted on whether to keep or repeal it, with critics saying that it is out of step with the originality standard. 3.4 What this suggests

2

Three positions are possible. One is the American view that purely machine-made output belongs to no one and falls into the public domain. Another is the British and Indian text, which gives the output to the person who arranges its creation. A third would give it to whoever invested in the system. The first is cleanest in doctrine, since it keeps human creativity at the centre. The second is friendlier to commerce but strays from the rationale for copyright.

4. Training data

4.1 The problem To work, a generative model is trained on very large collections of text, images, music or code. Much of this is copyrighted. Training involves making copies, at least temporarily. Rights-holders argue this is infringement on a vast scale. Developers argue that training extracts patterns and not expression, and that it should be treated as lawful, or as a permitted use. 4.2 The Indian statute India has no specific exception for text and data mining. The nearest provision is the fair dealing exception in section 52(1)(a), which allows dealing with a work for private or personal use including research, criticism or review, and reporting of current events.[6] Indian fair dealing is a closed list of purposes, unlike the open-ended American "fair use" doctrine. It is a difficult fit for commercial training at large scale, where the purpose is building a product. 4.3 The pending case The question is now before the Delhi High Court in ANI Media Pvt Ltd v OpenAI Inc, filed in late 2024, where a news agency alleges that its content was used to train a language model and reproduced in outputs. Publishers' and music-industry bodies have sought to join the case.[7] The litigation raises questions of jurisdiction, storage of copies, the reach of fair dealing and the status of outputs. Because the matter is pending, this paper takes no position on its outcome. 4.4 Other systems In the United States, trial courts have so far reached different results on fair use. Two federal decisions in 2025 found that training a model on lawfully obtained books could be fair use, while one of them also held that building a library from pirated copies was not covered.[8] An earlier decision involving a non-generative legal research tool found no fair use.[9] The picture is unsettled and appeals are pending. The European Union has an express text-and-data-mining exception in Directive (EU) 2019/790, with a commercial opt-out right for rights-holders who reserve their rights in a machine-readable way.[10] Japan allows training under a broad exception for information analysis, with limits where the use unreasonably harms the owner.

5. Reform options

Four ideas deserve attention. The first is a text-and-data-mining exception on the European model, allowing training on lawfully accessed works unless the owner opts out. It gives developers certainty. But an opt-out places the burden on creators, and small creators in India are not well placed to manage

3

machine-readable reservations. The second is a licensing model. A committee of the Department for Promotion of Industry and Internal Trade has reportedly considered a framework in which developers may train on lawfully accessed content but pay a statutory royalty collected and shared through a central body. This avoids the cost of a licence negotiation with every owner and gives creators a share, though it raises questions of rate-setting and how to divide income fairly. The third is a transparency duty. Developers would have to disclose the main datasets used for training. Without disclosure, a creator cannot know whether her work was used, and so cannot enforce any right. The fourth is clarity on outputs. Parliament could amend section 2(d)(vi) so that it applies only where a human has made a real creative contribution, defining what counts, and could state that purely machine-made content is outside protection or gets a short, sui generis term. The author favours a combination of the second, third and fourth. A paid licensing route offers creators something concrete and lets research and innovation continue. Transparency makes it enforceable. And a firm human-contribution test keeps copyright tied to creativity instead of investment.

6. Conclusion

The two questions of ownership and training should be answered separately. On outputs, Indian law gives the user a head start through section 2(d)(vi), but originality still demands human creative input, and the section needs updating to say so. On inputs, the absence of a mining exception and the narrowness of fair dealing leave developers and creators in a standoff that the Delhi High Court may begin to resolve. A system that pays creators, requires disclosure and keeps human creativity at the centre would give all sides something they can plan around. Waiting for courts to settle every detail case by case is possible but slow, and the technology will not wait.

References [1] Eastern Book Company v D.B. Modak (2008) 1 SCC 1. [2] Copyright Act 1957, ss 2(d)(vi), 13, 17, 22. [3] United States Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability (January 2025). [4] Thaler v Perlmutter, 130 F 4th 1039 (DC Cir 2025). [5] Copyright, Designs and Patents Act 1988 (UK), ss 9(3), 12(7), 178. [6] Copyright Act 1957, s 52(1)(a). [7] ANI Media Pvt Ltd v OpenAI Inc, Delhi High Court, CS(COMM) 1028/2024 (pending at the time of writing). [8] Bartz v Anthropic PBC (ND Cal 2025); Kadrey v Meta Platforms Inc (ND Cal 2025). [9] Thomson Reuters Enterprise Centre GmbH v Ross Intelligence Inc (D Del 2025). [10] Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market, art 4.

4

Cite as: Vikram, “Who Is the Author? Generative AI and Copyright in India”, *AICyberLex Nexus*, Vol. I, Issue I, Jul–Sep 2026, Article 5. Available at https://aicyberlexnexus.com
© 2026 AICyberLex Nexus. Open access under the terms stated on the journal website.