

Exploring the DPIIT's working paper on Generative AI and Copyright

On 8 December 2025, the Department for Promotion of Industry and Internal Trade's (DPIIT) committee (**Committee**) issued the Working Paper (**WP**) on Generative Artificial Intelligence (**AI**) and Copyright (Part I). The issue of AI training using copyrighted content and copyrightability of outputs were highlighted in critical policy documents, including the Principal Scientific Advisor's draft sub-committee's 2024 report (and later, the final AI governance framework from November 2025). These issues gained further prominence in November 2024, when Asian News International sued OpenAI before the Delhi High Court alleged copyright infringement.

Against this backdrop, in 2025, the DPIIT constituted a Committee, to identify key copyright issues arising from use of AI, examine whether India's copyright regime adequately addresses these issues; and provide recommendations. This Committee comprises of members from DPIIT, Ministry of Electronics and Information Technology (**MeitY**), NASSCOM, lawyers, and academia.

Basis a stakeholder consultation, the Committee has released this WP (Part-I) for public comments, open for consultation till 7 January 2026. Part-I focuses on the legalities of using copyrighted works for AI training and sets aside the question of copyright over AI outputs for Part II. It provides: (a) an overview of AI and current status of India's policy approach towards regulating AI; (b) key issues at the intersection of AI training and copyright law; (c) global approaches (for reference); (d) an assessment of various regulatory models; and a (e) proposed policy framework.

Here, we explore: (a) key AI and Copyright issues flagged in the WP; (b) DPIIT's assessment of regulatory models; (c) DPIIT's recommendations; and (d) the path ahead.

1. *AI and Copyright issues in India*

The WP discusses key issues under the current copyright framework, captured below:

- (a) **Question of infringement:** Section 14 of the Indian Copyright Act, 1957 (**Act**) grants a copyright owner certain exclusive rights, including reproduction, storage, translation, adaptation, and communication to the public. Any such activity without necessary authorisation amounts to infringement, under Section 51 of the Act. The WP details how the process (downloading and storing data to creating multiple temporary copies for model training) involves making reproductions of copyrighted works. While acknowledging industry view that training only extracts unprotectable patterns (non-expressive use), it underscores that the act of copying itself raises significant infringement concerns, creating a high-risk legal environment for AI developers. Ultimately, the WP acknowledges that the issue of infringement does not have a blanket answer, and is dependent on a case-by-case factual assessment.
- (b) **Fair dealing versus fair use exemption:** Section 52 of the Act provides "fair dealing" exceptions for specific purposes - private use, research, criticism, or review. The WP

emphasizes that this provision (i) is narrowly defined and purpose-specific; (ii) does not provide the broad flexibility of the U.S. "fair use" doctrine; and (iii) cannot be utilised without demonstrating how an activity fits the prescribed criteria. This narrow scope, combined with pending litigation, creates significant legal uncertainty, meaning businesses cannot confidently rely on the existing fair dealing defense to shield their AI training activities from infringement claims. Consequently, the WP concludes that amending or expanding the fair dealing provision is an inadequate and impractical solution. It argues that fair dealing is merely a legal defense, not a proactive right, and would always require a court to conduct a complex, case-by-case analysis to determine its applicability. This inherent unpredictability fails to provide the urgent legal certainty the AI industry needs to innovate.

2. *DPIIT's assessment of regulatory models*

The Committee assesses different regulatory models (discussed below); and concludes that traditional copyright models cannot effectively support AI training:

- (a) ***Text and Data Mining (TDM) exemption:*** India's copyright statute currently does not provide for a TDM exemption, which is generally enabled in other jurisdictions. However, a blanket TDM exception undermines copyright and offers no compensation or control to creators. An opt-out TDM regime is also unworkable, as it imposes heavy burdens on rightsholders, depends on full transparency from AI developers, risks degraded AI quality due to incomplete datasets. It also creates enforcement and compliance difficulties given the need for clear, interoperable, machine-readable notices.
- (b) ***Voluntary and direct licensing:*** The WP decisively rejects voluntary and direct licensing as an impractical and unworkable solution for AI training. The sheer scale of data required (of copyright holders) makes negotiating individual licenses logistically impossible. This model could create insurmountable entry barriers for startups and smaller innovators, effectively stifling competition and hindering the development of a robust AI ecosystem. If numerous rights holders refuse to license their content, it would create significant gaps in training data, leading to biased and less effective AI systems.
- (c) ***Collective Licensing (CL)/Extended Collective Licensing (ECL):*** CL and ECL offer a significant improvement over direct licensing by streamlining negotiations and reducing transaction costs. This model, managed by Collective Management Organisations (CMOs), provides a mechanism for handling consent and remuneration on a large scale. The WP recognizes that such systems are well-established in India and globally, presenting a viable structure for aggregating rights and simplifying access for AI developers. However, it concludes that a purely voluntary CL or ECL framework is inadequate for AI training because it creates critical business risks. The primary concern is the potential for CMOs to create "hold-out" situations, leveraging their collective power to demand excessive royalties or refuse licenses altogether, which would stifle innovation and create high entry barriers for startups.

- (d) **Statutory licensing:** The Committee champions statutory licensing as the most effective framework to resolve the AI training dilemma, viewing it as a powerful tool that balances innovation with creator rights. It argues this model provides AI developers with guaranteed, predictable access to the vast datasets they need, eliminating the risk of hold-outs and the impracticality of direct negotiations. By making data access a legal right tied to a mandatory payment obligation, statutory licensing offers the legal certainty the industry requires while ensuring creators are fairly compensated for their work. However, traditional statutory licensing is administratively burdensome for the massive scale of AI. To overcome this, it proposes an innovative "Hybrid Model" that modernizes the concept, captured in the next section.

3. **DPIIT's recommendation**

The Committee proposes a "Hybrid Model" that blends elements of India's statutory licensing regime with a more balanced framework for AI training. It acknowledges the legitimate concerns raised by unlicensed use of copyrighted content and notes that broad exceptions without consent would disproportionately benefit developers while eroding incentives for human creators. To sustain creativity and enable effective AI development, the Committee proposes a model that preserves broad data access while ensuring fair compensation for rightsholders.

Key features: This model proposes a mandatory blanket license with a statutory remuneration right for the creators and copyright holders for the use of all lawfully accessed copyright-protected works for AI training. Further, copyright holders cannot withhold their works from AI training. This approach is expected to (i) provide legal certainty for AI developers; (ii) ensures fair and inclusive remuneration for creators across both organised and unorganised sectors; (iii) reduces compliance and transaction burdens (particularly benefiting start-ups and smaller developers); (iv) and helps mitigate AI bias by enabling more representative and diverse training datasets.

Other aspects of the proposed working mechanism are captured below:

- (a) **Centralized Collecting Entity - Copyright Royalties Collective for AI Training (CRCAT):** This will operate as a government-designated, non-profit umbrella body formed by rights-holder associations under the Act. Its membership would consist of major copyright societies or Collective Management Organisations (**CMOs**), with one organisation representing each class of work to ensure broad, cross-sector representation. In sectors or regions where no CMO currently exists, the government will appoint nominees to serve on CRCAT's board until such bodies are established. It would be responsible for collection of royalties and disbursing them to member organisations.
- (b) **Royalty Rate Setting:** Recognising that traditional usage-based formulas are impractical, the Committee has proposed a flat revenue-share model in which royalties are calculated as a percentage of the global gross revenue generated by the commercialized AI system.

The model will apply retroactively, requiring developers who have previously trained or deployed AI systems using copyrighted works to pay royalties. This will be supervised by a government-appointed “Rate Setting Committee” composed of government, legal, industry, and technical experts, along with representatives from CRCAT and AI developers. It will set rates transparently, revise them every three years. The rates will also be subject to judicial review.

- (c) ***Distribution of Royalties:*** Under the proposed framework, every AI developer must submit an AI Training Data Disclosure Form to CRCAT, providing a sufficiently detailed summary of the datasets used for training. Once an AI system is commercialised, developers must annually calculate and remit royalties to CRCAT based on the flat revenue-share rates set by the Rate Setting Committee. CRCAT will then allocate the total royalty pool across different classes of works according to the proportion of their use disclosed in the developer’s form.

Each CRCAT member will maintain a Works Database for AI Training Royalties, where any copyright owner may register their works. CRCAT members distribute the allocated share to registrants according to their own AI Training Royalty Distribution Policies, using either pro-rata or value-based models. For sectors without an existing CMO, CRCAT temporarily holds royalties for up to three years, which can be claimed retrospectively if a CMO is established and becomes a member of CRCAT. However, if no CMO emerges in that period, the unclaimed amounts are transferred to a welfare fund for rightsholders in those sectors. Only registered works are eligible to receive royalties.

- (d) ***Grievance Redressal and Monitoring:*** CRCAT and each of its member will maintain a grievance redressal mechanism with a designated cell to address issues such as royalty non-payment and disputed ownership claims. These grievances must be resolved within timelines, and all decisions will remain subject to judicial review.
- (e) ***Burden of Proof and Exclusion of Proprietary Data:*** If an AI developer is sued for not paying the mandated royalties; and denies using third-party content for training, the burden of proof will rest on the developer. The law will presume the rightsholder’s claim is valid unless the developer can demonstrate compliance and prove that its system was not trained on the disputed content.
- (f) ***Injunctive Relief:*** The Committee concludes that introducing a statutory bar on injunctive relief would be premature. Instead, it recommends retaining the current judicial approach, giving courts the discretion to balance the interests of AI developers and copyright owners.

4. The path ahead

The WP proposes a fundamental market recalibration for India’s AI ecosystem, with distinct trade-offs for all stakeholders. For AI developers, both large and small, the framework offers a significant advantage by providing legal certainty and permission-free access to vast datasets.

This could lower the risk of litigation and level the playing field for startups. However, this comes at the direct cost of a new, mandatory revenue-sharing obligation. This will impact the business models of all commercial AI ventures.

Conversely, for content creators and rights holders, the model establishes a crucial and guaranteed new revenue stream, formally recognizing the value of their work in the AI value chain. This financial benefit is balanced by the loss of their right to withhold content. Ultimately, the Committee seeks to stabilize the ecosystem by replacing legal ambiguity with a predictable and regulated framework, which balances innovation incentives with creator compensation.