![Takshashila Institution logo] TAKSHASHILA INSTITUTION

Takshashila Policy Advisory - Working Paper on Generative AI and Copyright (Part 1)

Bharath Reddy

Version 1.0, February 2026

The working paper on Generative AI and Copyright aims to level the playing field for start-ups and compensate creators for their data being used for training AI. While the goal is noble, the heavy-handed approach threatens to stall innovation and ultimately fails the very creators it seeks to protect. India should adopt a market-driven approach that incorporates a text and data mining (TDM) opt-out mechanism for publicly accessible data and support commercial licensing agreements for proprietary datasets.

Contents

The working paper "One Nation, One License, One Payment: Balancing AI Innovation and Copyright"[1], published by the Department for Promotion of Industry and Internal Trade (DPIIT) committee, aims to level the playing field for start-ups and compensate creators for their data being used for training AI. While the goal is noble, the heavy-handed approach threatens to stall innovation and ultimately fails the very creators it seeks to protect. India should adopt a market-driven approach that incorporates a text and data mining (TDM) opt-out mechanism for publicly accessible data and support commercial licensing agreements for proprietary datasets.

The feedback on the working paper is presented below in two sections. The first section examines the potential outcomes if the recommendations outlined in the Working Paper are implemented. This analysis aids in anticipating the unintended and mitigating undesirable outcomes or enhancing desirable ones. The second section outlines some recommendations to achieve the objectives outlined in the paper.

1   Projected Outcomes

1. Impediment to Indian Innovation: The proposed framework will likely have significant negative implications for India's domestic AI industry and may drive companies abroad.

   Rationale:

   - The requirement to share a percentage of "global revenue" may be viewed as a penalty on Indian startups attempting to scale and serve a global market. Faced with such costs, Indian startups are likely to move to jurisdictions with friendlier copyright laws, such as Singapore or the US.

   - AI model developers might choose to exit the Indian market, leading to less competition in the market. Many global AI labs are investing in capturing a share of the market without a visible path to profitability. The proposed framework also requires model developers to disclose training datasets in detail. Barring a few open AI models, companies rarely disclose this data, as it is an ingredient of their secret recipe or moat. Forcing these measures would slow the pace of innovation and adoption in India to a suboptimal state for creators, developers and users alike.

     There is a precedent for this. When Canada passed a law forcing big tech to pay news publishers for linking to their content, the platforms just turned off the service, impacting discoverability and access to the content.

   - One of the stated objectives of the proposed framework is to level the playing field for all, including start-ups

and small players.   If the proposed framework is implemented, it is likely to benefit Indian AI model developers as it provides access to data through mandatory licenses while imposing costs on larger models, which might have significant revenues. However, downstream AI application developers (where India has a significant presence) will be limited in their choice of AI models.

- The AI innovation ecosystem consists not just of generative AI model developers, but also downstream application developers. Competition and choice across different stages of the AI value chain are essential for Indian companies to thrive on the global stage. If global model developers choose not to make their models available in India owing to onerous copyright requirements as outlined in the draft policy, developers will have limited choice in choosing the best model that fits their needs.

- India stands to gain significant productivity gains through the diffusion of AI. These benefits come from building applications that solve for India even if the underlying models are not Indian. This is the gradual process where AI applications are adopted across various sectors of the economy. In a bid to address compensation for creators, competition across the ecosystem and the pace of adoption of downstream applications will suffer.

2. Administrative Collapse and "License Raj": The administrative infrastructure required to implement the proposed revenue collection and distribution framework is unduly burdensome for firms and require massive state capacity to implement, making it likely to be ineffective.

   Rationale:

   - The Copyright Royalties Collective for AI Training (CRCAT) would be tasked with collecting and distributing royalties for billions of data points. Verifying these owners (KYC, banking details) will be practically impossible, making the system unwieldy to manage.

   - A fair rate that represents the value of Indian datasets in the overall training data and who gets what share of the revenue are very difficult issues to address. The idea that a single government-mandated body can determine an appropriate royalty rate (for AI model developers) and a revenue split (for content creators) goes against the core economic tenet of price discovery through the market mechanism that can produce high value for all stakeholders.

   - Compliance requirements to participate in the proposed framework might also be a deterrent for many genuine

content creators and publishers.

- There is a likelihood that royalties may not reach creators. Meghna Bal mentions the poor track record of existing welfare boards [2] (₹5,200 crore in a Labour Welfare Fund) where funds are collected but not disbursed.

3. CRCAT Becomes a Monopsony Entity: The proposed centralised CRCAT as a mandatory monopsony for Indian-sourced content for training forecloses the possibility of a competitive market.

Rationale:

- In various parts of the world, there are entities that have emerged that specifically collect data from data owners and license such data to AI model developers. With competition among such entities, data owners can benefit by participating in a "copyright pool" that maximises their revenue.

- Such collectives can also assist small content creators in bridging the technical awareness to implement machine-readable opt-outs (e.g., robots.txt) that the DPIIT report states as a challenge.

4. Failure to Benefit Creators: Despite compensation for content creators being a stated goal, the benefits to creators will be minimal and skewed towards larger aggregators. [3]

Rationale:

- Rahul Matthan observes[4] that much of the licence revenue will end up in a 'black box,' from which most individual artists will not be able to claim a share. Only established artists whose works are overrepresented in the dataset will be able to establish a claim. The proposed revenue calculations can be unrelated to the value of the content for model training. He compares this to a "legally sanctioned transfer of wealth from small individual creators to established artists and the intermediaries that serve them". To understand how this fails individual creators, we see an example in the music streaming industry, where royalties from streaming providers fail to add up to a living wage for most artists.

- Collective management organisations are unlikely to represent the vast diversity of independent artists and creators. By removing the opt-out mechanism, the framework effectively denies creators the right to control and monetise their work on their terms. Independent artists and creators are bound by the terms decided by a collective where they are not adequately represented.

5. Legal Uncertainty and Litigation: The implementation of these recommendations is likely to trigger a wave of legal challenges. [5] [6]

    Rationale:

    - By mandating license requirements, the policy actually misses the main point of copyright. The goal of copyright law is to promote human creativity by offering protection so that "copycats" cannot steal with impunity. However, AI models are not copycats - while data is used for training, AI models are not direct competitors and do not necessarily reduce the market for specific pieces of copyrighted content. While AI-generated content competes as a whole with human content, it is non-specific to a particular creator, unless there is style mimicking.

    - The rates for collection and distribution of royalties are not an exact science and can be challenged legally. Generative AI systems are probabilistic engines, not databases. They work by predicting tokens based on billions of parameters tuned during training. It is impossible to say with certainty if, and by how much, a specific work contributed to a specific output.

    - Ambika Agarwal points out[7] that the only feasible opt-out option would be for rights holders to go offline, which is not the intent of our copyright laws. She also notes that the mandatory license conflicts with voluntary open-source licenses (like Creative Commons), potentially overriding the rights holder's intent to allow free distribution.

6. Creation of Biased Models: The proposed framework could lead to models that perform poorly in understanding Indian cultural or social nuances.

    Rationale:

    - Global AI model developers may choose not to train on Indian datasets to avoid liability. This creates a scenario where AI systems are powerful generally, but perform poorly when addressing India-specific issues or understanding Indian cultural or social nuances. We risk importing models that "don't understand us".

2   Recommendations

1. India should adopt a purpose-neutral Text and Data Mining (TDM) exception for both commercial and non commercial purposes, conditional on a machine readable opt-out at the point of availability. This approach, recommended by Nasscom in their written submissions to DPIIT, balances

the need for rapid AI training with sovereignty for content creators.

For content that is not available in the public domain or where the creator has explicitly opted-out of TDM, a market for data should facilitate price discovery and lower transaction costs. Unlike the CRCAT recommended by DPIIT, which risks becoming a monopsony player with the power to distort prices, market based data cooperatives can pool data from different rights-holders and negotiate competitive, value-based licensing contracts with AI developers.

Rationale:

- TDM exception overcomes the entry barriers created by mandatory upfront licensing, which disproportionately hurts start-ups. A machine-readable opt-out (e.g., robots.txt) balances innovation with the right of the creator to reserve their works.

- The DPIIT Committee has criticised opt-outs for being burdensome to individual creators. Data Cooperatives solve this through aggregation. By acting as intermediaries, these cooperatives bridge the technical literacy gap, managing opt-outs and licensing at scale, functioning as a market-led alternative to a centralised state mandate.

- To ensure legal certainty, AI developers must be required to maintain rigorous data provenance records. Under such a system, the burden of proof for infringement would lie with the rights-holders or their representative cooperatives, who are better equipped than individuals to pursue litigation in cases of non-compliance.

- Unlike the flat rates that can distort markets and cause deadweight loss, a market-based mechanism allows for dynamic pricing based on the actual utility of the data to specific AI models.

2. The technical and operational capacity of the Competition Commission of India (CCI) and the judicial system should be expanded to lead to faster resolution of anti-competitive data practices and copyright abuses. Rather than creating the CRCAT, the recommendations in the CCI's market study on AI[8] provide more robust pathways.

Rationale:

- Given the scale, complexity and pace of AI development, regulatory capacity building is essential. This requires the development of specialised expertise in AI technologies, data science, and computational methods, and tracking global regulatory developments.

- The study also recommends self audits, frameworks to improve transparency, inter-regulatory coordination

and international cooperation which can help regulation
keep pace with innovation.

Endnotes

1. Department for Promotion of Industry and Internal Trade (DPIIT), "One Nation, One License, One Payment: Balancing AI Innovation and Copyright (Working Paper, Part 1)," Ministry of Commerce & Industry, December 8, 2025, Link

2. Meghna Bal, "India is trying to fix the copyright issue with AI training. It's doing it wrong," The Indian Express, December 16, 2025, Link.

3. Swaraj Paul Barooah, "Bulldozer Justice Comes to the Digital Sphere: Looking at the GenAI-Copyright Report," SpicyIP, December 25, 2025, Link.

4. Rahul Matthan, "Reverse Robin Hood," Ex Machina, December 18, 2025, Link.

5. Prashant Reddy T., "AI copyright, dead on arrival?," The Economic Times, December 19, 2025, Link.

6. Akshat Agrawal, "Why India's DPIIT proposal on AI training fees is a faith-based expansion of copyright law," ETGovernment, December 22, 2025, Link.

7. Ambika Aggarwal, "One Nation, Forced Licenses, Multiple Payments: (Un)Balancing AI Innovation & Copyright," SpicyIP, January 4, 2026, Link.

8. Competition Commission of India (CCI), Market Study on Artificial Intelligence and Competition, September 2025, Link.

TAKSHASHILA
INSTITUTION

The Takshashila Institution is an independent centre for research and education in public policy. It is a non-partisan, non-profit organisation that advocates the values of freedom, openness, tolerance, pluralism, and responsible citizenship. It seeks to transform India through better public policies, bridging the governance gap by developing better public servants, civil society leaders, professionals, and informed citizens.

Takshashila creates change by connecting good people, to good ideas and good networks. It produces independent policy research in a number of areas of governance, it grooms civic leaders through its online education programmes and engages in public discourse through its publications and digital media.