Journal of Intellectual Property Studies Vol. X (1), May 2026, pp. 1-18

MANDATORY BLANKET LICENSING MODEL ON GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT: A CRITICAL ANALYSIS OF THE DPIIT PROPOSAL

DR NARENDRAN THIRUTHY* & MR. DEBASHISH DASH**

ABSTRACT

*India's aspiration to nurture Artificial Intelligence ["**AI**"] developers, apart from being an AI consumer, is reflected in its IndiaAI mission that has started in March, 2024. To realise this potential, a DPIIT committee was formed in early 2025 to look into the copyright roadblocks for the adoption of AI, and the committee has recently released a working paper, suggesting sweeping changes to Indian copyright law to accommodate the transforming technology. After deliberations of all possible solutions to AI's copyright problem, the paper has suggested a hybrid model, integrating the underlying principles of both statutory exceptions and licensing regimes. The committee rejected the industry body NASSCOM's proposal for introducing a text and data mining ["**TDM**"] exception like that of the European Union, and favoured a mandatory licensing model, wherein the data owners cannot withhold their data from being ingested into the AI training process. This compulsory taking of data is complemented by the right of the content holder to receive fair remuneration from the AI companies. While the model is certainly a positive development in terms of bringing the rightsholders to the centre of AI – copyright discussion, there are many theoretical, tactical and practical challenges to the suggested model from a law and economics point of view. The issues in this model are manifold. First, it raises incompatibility with India's international treaty obligations. Second, there are apprehensions about the fixing of prices and their probable distortionary effect on the market. The next issue comes in terms of the sharing of royalties as a portion of the global revenue of AI companies. And lastly, the difficulties around the implementation of the model and its retrospective effect. All these issues are critically discussed in this article, besides appending suggestions to make it more acceptable and sustainable from a copyright perspective.*

I. INTRODUCTION

AI has become an integral part of our common lexicon, influencing daily conversations, debates and discussions, ever since the introduction of market AIs, like ChatGPT, in late 2022. Unlike any other technologies that emerge and fade in our collective consciousness, the euphoria around AI and its impact is refusing to subside. On the contrary, it is being touted as one of the most significant inventions of human history, and the interest around the subject is growing remarkably with each passing year. While some experts believe that AI will herald a new era of positive

* Dr Narendran Thiruthy is an Assistant Professor at the Rajiv Gandhi School of Intellectual Property Law, Indian Institute of Technology Kharagpur.
** Mr. Debashish Dash is a PhD Scholar and a Junior Research Fellow at the Rajiv Gandhi School of Intellectual Property Law, Indian Institute of Technology Kharagpur.

transformation across society and economy, there is an equal amount of hesitation and skepticism towards its adoption, primarily due to a lack of transparency, ethical concerns, and disruptive potential.[1] The current phase of AI growth has been propelled by its promise to solve many of our problems across industrial process automation, healthcare, education, disaster response, planning, governance, and more. The AI wave is continuing to disrupt many fields, and it will surely rewrite the future of our planet, hopefully in a positive way. Nonetheless, the growth of AI has also, knowingly or inadvertently, stepped into many legal minefields, springing up challenges around data privacy,[2] intellectual property (IP), liability issues,[3] etc. Among the IP issues, the conflicts around copyright law have emerged as the most vexatious and far-reaching. Inevitably, it has generated the most heated discussions, both in the courtroom and outside, in an attempt to settle the issues.

The AI's conundrum with copyright is inherent in the technology itself, as it is heavily dependent on diverse datasets that are often copyrighted, on whose foundation the AI model is trained and tested. Large language models ["**LLMs**"] derive intelligence from the underlying data and function as an obedient learner, only to produce analysis and inference efficiently, as and when required. The current capabilities of AI have been made possible by several years of advancement in data mining – the technology through which large amounts of data are analysed to extract patterns, trends, and correlations.[4] Since data includes both text and non-textual data, the technology is also referred to as TDM. While the data that has fallen into the public domain may have a different standing in the debate, the use of copyrighted data, which forms a major part of the training dataset, heavily impinges upon the rightsholders, by curtailing the method of exploitation of their economic rights conferred by copyright law. Prima facie, the extraction of data defeats the right of reproduction; the processing of data strikes the right to make an adaptation; AI outputs containing expressive elements of the work violate the right to distribution and public communication, among others.[5] Moreover, it is argued that the TDM life cycle also infringes upon the authors' moral rights by tampering with attribution and distorting the integrity of the original

[1] Ray Eitel-Porter, *Beyond the Promise: Implementing Ethical AI*, 1 AI ETHICS J. 73 (2021).
[2] *Rethinking Privacy in the AI Era: Policy Provocations for a Data-Centric World*, STANFORD HAI (Feb. 22, 2024), https://hai.stanford.edu/policy/white-paper-rethinking-privacy-ai-era-policy-provocations-data-centric-world.
[3] Herbert Zech, *Liability for AI: Public Policy Considerations*, 22 ERA FORUM 147, 150 (2021).
[4] (Data mining is a tool to uncover patterns and insights from large data sets using computational and statistical methods. In the legal understanding, one definition from the European Union (EU) can be culled out for reference. As per Article 2(2) of the 2019 Digital Single Market Directive of the EU, Text and Data Mining (TDM) is defined as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations". Though AI is an application of TDM technology, both are fuelling and complementing each other).
[5] Joseph Will, *Rage against the Machine: Copyright Infringement in AI-Generated Music Note*, 31 J. INTELL. PROP. L. 378, 393 (2024).

work.[6] Understandably, the rightsholders have challenged the use of their work for AI in various courts all around the globe, major ones being the case of the *New York Times v. OpenAI* ["**NY Times case**"][7] *& Microsoft* in the USA and *ANI v. OpenAI* ["**ANI v. OpenAI**"]in India.[8] While the outcomes of these two cases are still pending before the respective courts, a significant legislative effort is also being made in the respective countries to address the perplexing questions surrounding AI's copyright issues.

The U.S. Copyright Office is leading the effort through a series of reports on copyright and AI,[9] suggesting areas where legislative reform may be required while discussing the evolving jurisprudence of AI's impact on copyright law. The most recent report (third one in the series) arrived in May, 2025, titled "Generative AI Training," wherein the use of copyrighted content in AI training is scrutinized to ascertain its legal tenability. The report endorsed the view that the use of copyrighted material for AI training is essentially an act of infringement, and opined that it would be problematic to outrightly declare it as fair use.[10] As a solution, it suggested licensing in case of commercial use to mitigate risk, though it did not favour compulsory licensing of AI training data.[11] Significantly, it cautioned against any kind of hasty state regulation and instead recommended market-based solutions to these highly nuanced issues. Similarly, India has entered the AI discussion, with the publishing of the first report from the DPIIT, in Dec, 2025.[12] The DPIIT committee was formed in May, 2025, to look into the intersection between copyright law

[6] Joshua Yuvaraj, *I Do Not Consent: Moral Rights and Machine Learning in Australian Copyright Law*, SSRN (Oct. 7, 2025), https://papers.ssrn.com/abstract=5576350.

[7] (The case of New York times (*The New York Times Company v. Microsoft Corporation*, 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023) is an ongoing litigation, and it has been filed in the United States District Court for the Southern District of New York. This was one of the first few cases in the United States wherein the AI's copyright problem was brought into the judiciary, interestingly, after a failed negotiation between the parties. As per the latest information, the judge has allowed many of the copyright infringement claims to proceed further, rejecting the motions of OpenAI to dismiss the same. The outcome of this case will bring further clarity on the US position on AI training).

[8] (The case of *ANI Media Pvt. Ltd. v. OpenAI OPCO LLC & Anr.* (CS(COMM) 1028/2024), has been filed in the High Court of Delhi in 2024. This is also a pending litigation and first of its kind in India to decide on the copyright infringement in AI training in the context of copyright law. While OpenAI has agreed to block ANI content from its future training, there are many legal questions that would be answered by this case).

[9] (Since 2023, the U.S. Copyright Office has been examining the impact of AI on copyright law and policy. As of December 2025, it has published three reports in this series; first one on digital replicas, second one on the issues of copyrightability and third one on the Generative AI training issues. More information *available at* https://www.copyright.gov/ai/).

[10] (The official report of the U.S. Copyright Office (p. 107) (Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication Version, May, 2025), concludes by stating that "various uses of copyrighted works are likely to be transformative... but, making commercial use of copyrighted works to create content that can compete in the existing markets ...goes beyond established fair use boundaries").

[11] (As stated in the USA Copyright Office's May 2025 report (p.104), a compulsory licensing regime for AI training will have "significant disadvantages", owing to the fact that they will tend to make royalty rates inflexible, and after their adoption in the industry, it would also be tough to undo those changes. Further, it will thwart the possibility of developing market-based solutions).

[12] *DPIIT Publishes First Part of Working Paper on AI–Copyright Interface*, PRESS INFORMATION BUREAU (Dec. 09, 2025), https://www.pib.gov.in/PressReleasePage.aspx?PRID=2200741&reg=3&lang=2.

and AI, and suggest solutions to the problems posed by their interaction. As per the mandate, it has been published as a "*Working Paper on Generative AI and Copyright,*" suggesting many changes in the current copyright law, and most importantly, the introduction of a hybrid licensing model, as a "one nation one license one payment" model. The model indeed takes away the power of the copyright owner to deny licensing of their data, with a guarantee for compensation. The proposal has gathered mixed reactions from the stakeholders, with the industry body NASSCOM disagreeing with the idea of such licensing, mainly at the prospect of raising the cost of compliance, besides being difficult to implement (if not impossible).[13] The idea of a hybrid model to provide copyrighted work to train AI, through a right based jurisprudence, along with a promise of fair compensation, is a new frontier in the discussion, and it will have sweeping implications in the domain of copyright law. The desirability of such a model, especially from the standpoint of the primordial copyright objective of fostering creativity, is, arguably, a point of contention, in addition to the practicality of implementing such a system. Therefore, a detailed jurisprudential analysis is needed on this proposal, including its implication to our society, from the law and economics point of view.

The article first discusses the core substance of the model, as it is proposed by the DPIIT. Then, it dives into segment wise assessment of the model. In the assessment, it makes separate analysis into the effects the model brings with respect to the treaty obligations India have, followed by jurisprudential analysis, pricing issues and its effect on the market, operational difficulties in implementing the model, and the issue of retroactivity in application, among others. Lastly, the article concludes by making few suggestions that can remedy the difficulties posed by the model.

II. THE PROPOSED MODEL IN BRIEF

The proposed model, conceptualized as a mandatory licensing model, has been necessitated to resolve the impasse between access to AI training data and the obligation to compensate the rightsholders. As a hybrid approach, it has prudently stitched the ideas behind statutory exception (like fair use) and licensing regime, by combining both the right to access and right to compensation aspects with respect to the AI training data. The model proposes that the AI developers shall have the right to ingest all the lawfully accessed data for AI training, and no rightsholders can withhold their works from such a purpose. In the same breath, it has bestowed copyright holders with the right to obtain fair compensation for their work used in the training process. The payment mechanism is suggested to accrue only after the commercialization of the

[13] Sejal Sharma, *Nasscom Pushes Back against DPIIT Plan*, HINDUSTAN TIMES (Dec. 11, 2025), https://www.hindustantimes.com/india-news/nasscom-pushes-back-against-dpiit-plan-101765436303176.html.

model, which is intended to provide relief to the startups who are going through the critical designing and initial testing phase. The royalty is suggested to be distributed as a percentage of the global revenue of the AI company and the rates are slated to be decided by a government-appointed rate setting committee. As per the DPIIT working paper, the proposed rate setting committee shall consist of senior government officers, and several experts from legal, financial, and technical domains, having knowledge of emerging technologies. The said committee will be expected to set royalty rates based upon the expert recommendations, and review the rates every third year to cater the changing needs of market and technology. Moreover, the royalty rate will be subject to judicial review. The model suggests for a retroactive application of the licensing model, in order to undo the damages already occurred and provide a level playing field to all the players.

Additionally, it suggests the creation of an institutional architecture to implement the model, through a centralized entity in the name of Copyright Royalties Collective for AI Training ["**CRCAT**"]. Further, it mandates the AI companies to furnish an AI Training Data Disclosure Form, delineating the dataset use in their AI system, to bring the much-needed transparency and accountability in their working. The disclosure form shall contain a "Sufficiently Detailed Summary" of the dataset, and shall include the "category" (and subcategory) of data as per Section 14 of the Copyright Act of 1957, mentioning the classes of work like text, images, videos files etc. Additionally, the "sources" of the data, such as social media, online libraries, proprietary dataset etc., shall also be mentioned, along with the "nature" of the data, such as news, literature etc. For the rightsholders, a system of registration is prescribed to establish their connection to the governing apparatus that will implement the hybrid model and distribute royalties accurately. Finally, the report has also suggested a value-based matrix to track a class of work and its contribution to the neural network, which can help in determining the royalty rates. To implement all these suggestions, the DPIIT paper has proposed to bring necessary amendments in the Indian Copyright Act of 1957 and the Copyright Rules thereunder.

III. ASSESSMENT OF THE MODEL

The following discussion will examine the model's desirability and suitability in the Indian context, focusing on its strengths and weaknesses from a copyright perspective. The theoretical basis of each component of the model will be examined, keeping in mind the larger objective of copyright law, and its correlation with the general IP jurisprudence. Further, the operational feasibility of the model will be discussed, considering the Indian experience, and also drawing from other jurisdictions, as the technology that is working behind an AI system is largely jurisdiction agnostic, giving similar experiences in other jurisdictions.

A. Positive Shift in The AI Copyright Discourse

In the last few decades, the rapid advancement of technology has constantly threatened the domain of copyright law. The effect is particularly intense in the case of new and emerging technologies in the digital world, which have been disturbing the existing copyright markets by simplifying the access and copying of protected works with unmatched scale and anonymity. Historically, copyright law has responded to technological challenges by broadening its scope through judicial interpretation when required and by adding new rules to address them.[14] The constant disruptions that were brought by new technologies have forced the copyright law to evolve and recalibrate in order to remain relevant and effective. However, it is not always desirable to unnecessarily expand the interpretation of fair use to embrace newer technologies. In this context, the DPIIT proposal, which has centred the discussion around the rights of the content owners, is a welcome step. Instead of taking a pro-technology stance to deem the copying of content for AI training as fair use, it has clearly and unequivocally discarded the argument of fair use, and accordingly, suggested methods to compensate the right holders. Further, it has made the right to be compensated a central issue and has deliberated on various ways to design a suitable framework for its implementation. Therefore, this is a positive development: the discourse around AI's copyright problem is shifting from unreasonably accommodating the technology to acknowledging the legitimate interests of rightsholders, economic or otherwise, and coalescing policy development around the protection of those rights. This is not to say that it is advocating for an expanded copyright, this seems to be an effort to balance the urgent technological aspirations of society with the need to incentivise the creative community of society. Exploiting the copyrighted work for AI training or TDM, without finding a suitable mechanism to compensate the rightsholders, will negatively impact the sustainability of the same, besides ensuing numerous litigations in this space that would threaten the domain with a lot of risk and uncertainty.

B. Issues with International Treaty Compliance

IP law is fundamentally territorial, applying within a country's jurisdiction. However, international treaties also play a significant role in the administration of IP, by prescribing minimum standards of protection, apart from establishing global frameworks for harmonizing the law, and its procedure, to ensure international co-operation. In the case of copyright, the Berne Convention of 1886 ["**Berne Convention**"] and the Trade-Related Aspects of Intellectual Property Rights

[14] (With the progress of technology came the need to reshape the contours of copyright law. In the Indian context, computer programs and databases were added to the definition of "literary work" (*see* section 2(o) of the Copyright Act of 1957) to ensure their protection under IP law. Similarly, the regulations around technological protection measures were also introduced over time, and many more changes followed).

["**TRIPS**"], 1995, have been the foundational treaties, and India is a signatory to both. Consequently, India is obliged to honour the treaty obligations and adhere to the minimum standards prescribed in those treaties. In this context, the DPIIT proposed licensing model, perhaps, will come into conflict with the famous "three-step test" of the Berne Convention that has become the gold standard for checking the acceptability of copyright exceptions. The three-step test permits limitations and exceptions to an author's rights, provided that, they are meant for certain special cases and do not conflict with the normal exploitation of the work, besides also ensuring that the exception do not unreasonably prejudice the author's legitimate interests. This test, which later got incorporated into Article 13 of the TRIPS agreement, has three facets: first, the exceptions must be narrow in scope and have clear definition; second, it should not deprive the rightsholders of actual or potential source of income; and third, it should not create any unreasonable harm to the rightsholders. Though Indian Copyright law regime has exhaustive list of copyright exception in terms of fair dealing, courts have used the tenets of the three-step test to decide matters, and the case of *Chancellor, Masters & Scholars of University of Oxford v. Rameshwari Photocopy Services* [Photocopy case] is one of them.[15] In this case, the High Court of Delhi ruled that 'educational use' is a special case, and when students use the photocopy of books for instruction, it doesn't conflict with the normal exploitation of the work, as students do not form part of the market for pricey books This test remains a bulwark against dilution of the author's rights, an important sphere within the larger framework of copyright, and prevents unwarranted application of copyright exceptions.[16] However, in the present context, the exception to the rights of content holders, suggested via the proposed DPIIT model, is considered through the lens of three-step test, not in a case specific scenario.[17]

When the rightsholders are stripped of the option to opt out their data from the AI training process, the limited power given by copyright law to control and prevent unauthorised access and reproduction of their work is further diluted. However, another factor in this equation is the right to receive remuneration. This kind of system upends the IP rights, converting it from a system of property rule to liability rule; thereby reducing absolute entitlement to only offering a right to claim compensation in case of harm (AI training in this case). To contrast property rule with liability rule, it is traditionally understood that, the former (property rule) allows a right to be taken away

[15] *The Chancellor, Masters & Scholars of the University of Oxford & Ors v. Rameshwari Photocopy Services & Anr*, 2016 SCC OnLine Del 5128.
[16] Daniel Gervais, *Making Copyright Whole: A Principled Approach to Copyright Exceptions and Limitations*, 5 U. OTT. L. TECH. J. 1, 9 (2008).
[17] Rubén Calderón, *AI Training Through Copyrighted Works as Infringement: Perspectives Under the Berne Three-Step Test and the Pane Fair Use Test and Plausible Solutions*, 65 IDEA (2024).

from the owner only through a voluntary transaction, and at a price that is agreed by the owner. However, in liability rule, the entitlement can be taken without the owner's consent, provided there is payment of damages.[18] While it is true that IP rights are not absolute rights like tangible property, and are invariably laced with limitations and exceptions, the choice of making economic exploitation, or not, is particularly relevant from the rightsholder's perspective. On the three-step test, the first factor is going against the proposed model, as it fails to make a special case due to its expansive breadth that covers all lawfully accessed works for AI training. The case of blanket license covering all the available work is, arguably, a general derogation of copyright, rather than a special one. Therefore, the first factor of the three-step test will disfavor the proposed model. One way to remedy this situation could be to add a "conditional opt-out" mechanism, wherein, rightsholder can withdraw their work, and thereby, preventing the licensing regime becoming an absolute general rule. Permitting authors to opt out in case of probable injury to reputation can be one of them. Similarly, in case of commercial training, if the rightsholders can present an alternate feasible licensing mechanism, they can be allowed to withdraw from the mandatory regime. However, such a system could also complicate the regime, and therefore, should be approached with caution.

As far as the second factor in terms of the conflict with the normal exploitation is concerned, the situation is fast evolving. The swift growth and expansion of the AI training data market is going to change the dynamics of judging this factor, for data licensing can become the new normal, with umpteen number of publishing companies already signing deals with AI entities to permit their data for AI training. Another bone of contention in this factor is the difficulty in reconciling with the one nation one license in a government decided rate, effectively nationalizing the private market, and thereby stripping the rightsholders their power of negotiation in a free market. Additionally, owing to the uniformity of this licensing regime, few authors will feel undercompensated, asserting that their work could have gotten a better deal, while other may feel overcompensated for their work. In a state regulated regime, there will surely be a difference between the market price of the data and the statutory price, hurting the market sentiment, and thereby, making it economically unviable in the long run. Thus, one nation one license regime will only impede the evolving of a licensing market, which could, otherwise, have become the new normal way of exploitation, given the demand of curated and organized AI trainable data.[19] In this scenario, the second factor in the three-step doctrine too shall go against the proposed hybrid

[18] Louis Kaplow & Steven Shavell, *Property Rules versus Liability Rules: An Economic Analysis*, 109 HARV. L. REV. 713, 757 (1996).
[19] David W. Opderbeck, *Copyright in AI Training Data: A Human-Centered Approach*, 76 OKLA. L. REV. 951 (2023).

model, given how the model may, in effect, prematurely meddle with the normal exploitation of the work.

On the final factor of weighing in the legitimate interests of the author, the acts of appropriating of the author's data without consent, although statutorily backed in this case, may unreasonably prejudice their short term and long-term interests of the rightsholders.[20] Not only that the exception to the author's rights be clearly defined as per the first factor, the economic impact of AI training must also be judged carefully to arrive at any conclusion on the Berne second and third factor.[21] Interests of authors go beyond immediate monetary compensation and includes the moral rights associated with the works. Though it is not clear whether there is indeed any violation of moral rights, the AI training process can certainly impinge upon moral rights at various stages of AI development, due to erasure of information on attribution, apart from fragmenting the integrity of the work.[22] Moreover, another significant erosion of legitimate interest will happen in the long run, in terms of the loss of control that is tied to the right to remuneration regime. Oftentimes, technology commercially replaces authors and creators over a period of time, and in this case, the AI training that occurs with their own works will eventually replace them, and they shall have no control over the same.[23] It is true that the market AIs are replacing the coders whose code have been used to training the same market Ais.[24] At least it shall obliterate most of the low skill jobs in the coding domain, as predicted by many experts.[25] Similarly, it will also be a reality that many of the creative professions will have their replacements in AI, as AI becomes more creative with incessant training on the foundation of plethora of similar works. In this context, the proposed model can become highly prejudicial to the interests of the authors. Therefore, after considering all the three factors, it would be safe to assume that the proposed DPIIT model will certainly go against the letter and spirit of the Berne three-step test, and will remain as an outlier, unless few balancing acts are carried out to empower the rightsholders.

---

[20] *Supra* note 18.
[21] Larissa Bersh, *An Exceptional Formality Under Berne: Evasion Of Copyright Protection Via The Eu's Text And Data Mining Exception*, 38 HARV. J. L. TECH. 337, 345 (2024).
[22] Rita Matulionyte, *Can AI Infringe Moral Rights of Authors and Should We Do Anything about It? An Australian Perspective*, 15 L. INNOV. TECH. 124, 5 (2023).
[23] Ben Zhao, *Replacement of Human Artists by AI Systems in Creative Industries*, UN TRADE & DEVELOPMENT (Mar. 28, 2024), https://unctad.org/news/replacement-human-artists-ai-systems-creative-industries.
[24] Casey Moffitt, *Is AI a Coder's Friend or Foe?*, ILLINOIS INSTITUTE OF TECHNOLOGY BLOG, https://www.iit.edu/blog/ai-and-coding-future.
[25] Mini Tejaswi, *AI Will Gobble up Most Low-Hanging Jobs of Coders,* THE HINDU (July 26, 2025), https://www.thehindu.com/business/ai-will-gobble-up-most-low-hanging-jobs-of-coders/article69859151.ece.

C. Unstable Theoretical and Jurisprudential Basis

The current legislative proposal has made few assumptions while laying the foundation for the mandatory licensing framework. While it may be true that the process of AI training violates copyright law, with a series of infringements over the training life cycle, the courts have delivered differing opinions in the same and the judicial decision making is still unsettled on this point in many jurisdictions.[26] In India too, the case of ANI has been pending for disposal, and therefore, the case of infringement has not been conclusively made out by any judicial decision with respect to the Indian copyright regime. In this situation, it may be argued that, the DPIIT incurred the risk of proposing a model prematurely, as, from a jurisprudential standpoint, suggesting a remedy before a wrong has been concretely established seems logically inverted. Moreover, the intention to tinker the copyright law in order to facilitate the availability of diverse datasets for AI training (and AI growth) may be undesirable. It talks about an easy and fair access to critical data for effective AI development and the role mandatory copyright licensing can play in facilitating the same. However, the intention of copyright law is not to remove roadblocks for the growth of technologies, much less for the data-driven and copy-reliant technologies like,[27] AI that will significantly exploit the underlying copyrighted work; rather the objective of copyright law is to encourage creativity and incentivize the authors.[28] On the contrary, the growth of an industry is the primary objective of an industrial policy, not IP policy; though their paths, sometimes, cross each other. Therefore, bending the copyright law to suit an industry seems rather unwarranted at this stage, when the market forces are actively trying to rejig the rules of the game, through negotiation and private legislation, and creating binding contracts between players to ensure predictability in the domain.[29]

Further, the objective to mitigate the risk of bias in AI and avoid hallucinations by ensuring the availability of training data seems like a misplaced priority from the perspective of the copyright

[26] (In the US, the case of *Thomson Reuters v. Ross Intelligence* (694 F. Supp. 3d 467 (D. Del. 2023), aff'd in part, summary judgment granted Feb. 11, 2025) was decided against Ross on the finding that the use of Westlaw's copyrighted headnotes in the process of AI training was not fair use. However, in *Bartz v. Anthropic* (PBC, No. 3:24-cv-05417, (N.D. Cal. June 23, 2025)), the AI developer was favoured on the ground that the copyrighted books used to train the LLMs was a "transformative" fair use. Similarly, in the case of *Kadrey v. Meta* (No. 3:23-cv-03417-VC, (N.D. Cal. June 25, 2025)), the act of copying books by Meta was held to be fair use, as it would not result in the dilution of the market. Therefore, owing to diverging opinions, the position in the U.S. is not settled. Other countries too have not reached a conclusion on the issue of fair use and the jury is still out on this point).

[27] (The term copy-reliant technology refers to the ones which copy massive amount of copyrighted work for various purposes. Examples include search engines, data mining tools, AI models, plagiarism detectors, software reverse engineering tools etc. This set of technology falls in fuzzy areas under copyright law, though few of them are accepted as necessary evil. More information *available at* https://lawecommons.luc.edu/facpubs/69/).

[28] Simone Schroff, *The Purpose of Copyright – Moving beyond the Theory*, 16 J. INTELL. PROP. L. PRAC. 1262, 1266 (2021).

[29] *AI and Content Licensing*, CREATIVE LICENSING INTERNATIONAL, https://creativelicensinginternational.com/licensing-brief/ai-and-content-licensing/.

law. IP law is founded on the principle of balance, which is sought between the society and the private individual or entity who hold the exclusive right to gain from such IP protection. Accordingly, the net gain to the society is the touchstone through which an IP policy gets tested. The bestowal of IP plays an important role in expanding the creative works available to the public and sustains their supply. Such a balance is missing from the proposed DPIIT framework. It assumes that the creative potential of AI, after training with a vast amount of copyrighted work, will be enjoyed by society as an equivalent to human creativity. However, AI as a creative force has not been promising, as yet, and it has been noticed that the AI creativity remains low quality and quite generic, incomparable to human ingenuity.[30] One more interesting aspect of IP policy is to ensure that copycat versions do not flood the market, and if they are detected, the law must punish them though infringement cases and other compensation mechanisms. However, when AI is trained through datasets comprising various copyrighted works, it is inevitable that the AI output will have some resemblance to the training works and will look like copies.[31] Therefore, the entire purpose of removing copycats from the market is defeated, and this may bring various undesirable consequences. Hence, more discussions on the theoretical aspects of the proposed framework must be carried out in order to identify the exact problem the policy wants to solve and map it with possible solutions to distinguish the best possible alternative.

D. Pricing Issues and Probable Market Distortion

The economic analysis of the proposed DPIIT framework, which has located pricing at the core of the suggestions, will play an influential role in determining the acceptability of the same. It is proposed that the royalty rate will be finalised by a central government committee, to be known as the rate-setting committee, consisting of government officers, and other experts from legal, financial and technical domains. To make it fair, the proposal envisages court intervention and judicial review of the rates. The report acknowledges the complexity of AI rate setting, owing to the black box nature of neural networks,[32] and therefore, shies away from suggesting any standard formulae of value assessment. Instead, it suggests a flat rate, as a percentage of the gross global revenue of the AI developer from the commercialisation of the AI that has been trained on such content. It avers that it is a better system than other available options, as there would be no need

[30] (Many technologists believe that the current wave of AI has not been exceptionally creative. Rather they call it "*AI slop*," which refers to flooding of low-quality, repetitive and nonsensical content generated by AI, only to clog and degrade the information world. More information *available at* https://theconversation.com/what-is-ai-slop-a-technologist-explains-this-new-and-largely-unwelcome-form-of-online-content-256554).

[31] Hiroaki Chiba-Okabe & Weijie J. Su, *Tackling Copyright Issues in AI Image Generation through Originality Estimation and Genericization*, 15 SCI. REP. 10621, 5 (2025).

[32] Davide Castelvecchi, *Can We Open the Black Box of AI?*, 538 NAT. NEWS 20 (2016).

for any upfront payment during training, and liability to pay will only accrue after the generation of revenue, something, it contends, will help the start-ups more than the large players in the AI developing market. Though it is not the primary objective of copyright law to remove barriers of technology, the report nonetheless asserts that the proposed model of waiving upfront payment will create a supportive environment of AI development, and benefit the public eventually, through the aftereffects of AI progress. While these suggestions may look good on paper, a deeper analysis of the individual components unsettles the domain from many angles.

The first issue comes at the introduction of global revenue into the rate-setting mapping. It is unclear as to why India would need to look for the global revenue of AI companies, as it would be punishing for AI companies that are looking for scale and multinational presence. Global revenue may look difficult to reconcile for those companies that are either not residents in India or do not have their place of effective management in India, though they may still do some business in India. The second issue revolves around a flat rate for all kinds of AI training data, without any hierarchy of the data, juxtaposed with a centralised process of decision-making that diminishes the market forces. Along with the progress of AI technology, a separate training data market is emerging, and in the absence of government intervention, the market will settle as per the demand and supply.[33] Many publishers have negotiated with AI developers to allow their data to be used in AI training,[34] and their agreements will be impacted by the proposed changes to copyright law, which would act as an overriding public policy. Moreover, future negotiations will be thwarted, and the market price of the data will not be factored in. Though the DPIIT report mentions considering market price, a government-decided rate will differ from the market price, and therefore, it may distort the AI training data sector. Further, a countrywide single and uniform rate for all types of data will be unable to factor in the variations in the quality of data, leading to the creation of a market in which specially curated data will get the same value vis-à-vis low quality or ordinarily curated data. This would certainly disincentivise the market forces and negatively impact the availability of quality data. As a corollary, it can be asserted that the model may bring unnecessary bureaucracy and inefficiency in the royalty setting domain, and instead of helping the AI sector, it may discourage the development of AI models in India, provided the aforesaid imperfections are not mitigated.

[33] Harriet Aryee, *Licensing AI Training Data: Legal Considerations and Key Contractual Clauses*, SSRN (Jan. 8, 2026), https://papers.ssrn.com/abstract=6045594.
[34] Bruce Barcott, *How the Emerging Market for AI Training Data Is Eroding Big Tech's 'Fair Use' Copyright Defense*, TECH POLICY PRESS (Mar. 4, 2025), https://techpolicy.press/how-the-emerging-market-for-ai-training-data-is-eroding-big-techs-fair-use-copyright-defense.

E. Operational Difficulties and A New Institution

In order to make the proposed framework a success, the question of attribution of a particular work and its AI revenue trail must be clearly ascertained. However, given the intricate nature of AI training, and the opaque nature of the learning process, it would, perhaps, pose the biggest problem in its implementation. The technology may itself fail to provide attribution with certainty if similar content is fed into the AI's training, and things will be particularly difficult for a certain type of unlearning model, which forgets the training data as a matter of policy, either to mitigate bias or to ensure regulatory compliance.[35] Further, lack of quality content for AI training often pushes the developers to rely on AI-generated data. However, "indiscriminate" use of such synthetic content for training causes incurable defects, resulting in the collapse of the model.[36] Nonetheless, even in these systems, the attribution would be difficult to be made out, due to loss of human content trail, and the resulting doubt on the provenance of content.[37] AI models work as probabilistic systems, predicting the best possible answer within a given context, such as filling in the blanks.[38] There are billions of parameters through which AI training happens, without specifically mapping the derived intelligence to the source data from which it is derived. Moreover, mention be made of the fact that, many times, the information about attribution is lost during the data pre-processing and data cleaning phases, which act as precursors to the training process. Therefore, on technical grounds, it seems impossible to accurately map the contribution of a single piece of copyrighted work in the AI system, and accordingly, it would be uncertain if a fair distribution of the revenue could possibly be made on this basis. Thus, the basis on which the payment of revenue is proposed is technically difficult to establish, at the very least, and it shall cause tremendous problems in real-time operation.

The second factor in this equation is identifying the beneficiaries who will receive the revenue. If a proper trail is maintained by the AI companies with respect to the work being fed into the AI training process, it would be easy to identify the eligible rightsholders. However, as of now, there is no legal obligation imposed on the AI developers to maintain such records, and in the absence of any obligations, there must be no such practice to maintain the records since the beginning of their AI training. Therefore, one positive suggestion that the DPIIT proposal has made is to make

---

[35] Guangyao Dou et al., *Avoiding Copyright Infringement via Large Language Model Unlearning*, *in* FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: NAACL 2025 5191 (Luis Chiruzzo, Alan Ritter, & Lu Wang eds.) (2025).

[36] Alice Gomstyn & Alexandra Jonker, *What Is Model Collapse?*, IBM (Oct. 14, 2024), https://www.ibm.com/think/topics/model-collapse.

[37] Ilia Shumailov et al., *AI Models Collapse When Trained on Recursively Generated Data*, 631 NATURE 755 (2024).

[38] Mitchell Hashimoto, *Vibing a Non-Trivial Ghostty Feature*, MITCHELL HASHIMOTO (Oct. 11, 2025), https://mitchellh.com/writing/non-trivial-vibing.

it mandatory to keep the record of training dataset, through an AI training data Disclosure Form. However, it may still be difficult to map all the rightsholders from the training data, due to errors and discrepancies, and it would have many imperfections. Nevertheless, DPIIT has proposed the setting up of a new institution, which will be charged with administering the royalty payment issues and implementing the hybrid model. The model has proposed the creation of an entity named and styled Copyright Royalties Collective for AI Training ["**CRCAT**"] and its associated bureaucracy, acting as a non-profit body. This institution is projected to act as a collective of collectives.

The centralised entity CRCAT will hold the responsibility of collecting the payments received by AI developers and ensuring the flow of royalties to the member organisations, who will, in turn, distribute the money to rightsholders. It is envisaged that only organisations can be members of CRCAT, and only one member will be allowed for a designated class of work. Members can either be copyright societies under Section 33 of the Copyright Act, 1957,[39] or be a collective management organization ["**CMO**"], set up by a class of rightsholders. The details of their eligibility and representation are proposed to be included as Copyright Rules. The establishment of all these structures will not only complicate the bureaucracy, but also bring a lot of government intervention, necessary or otherwise, to administer the royalty issues, ideally to be governed by market forces. They will bring compliance issues and create disputes on many occasions, such as, the cases when registration is cancelled or multiple organizations vie for being a member of CRCAT. Meanwhile, the governing board of CRCAT is designed to balance the representation, including the nominees from each member and also include representatives from sectors whose works are not covered by any organization. While this also may look good on paper, the history of the collective societies and their functioning casts doubt on the fairness in running of the institution.[40]

Another significant concern relates to the concept of lawful access to the data. It is not clear how lawful access is to be determined. Theoretically, how can the data be lawfully accessed if there is no permission or license to access the same? The DPIIT report mentions ensuring permission-free access to data for AI training, and in that case, it is appearing contradictory to think that

[39] (Section 33 of the Copyright Act of 1957 stipulates that only registered copyright societies in India will be able to carry on the business of issuing or granting licenses for copyrighted content, besides mentioning the process of registration. The section also allows the individual owners of copyrighted works to license their works, provided their obligations under the registered society are not in conflict with the same).

[40] Prashant Reddy T, *AI Copyright, Dead on Arrival?*, THE ECONOMIC TIMES, https://economictimes.indiatimes.com/opinion/et-commentary/ai-copyright-dead-on-arrival/articleshow/126082127.cms?from=mdr.

permission free access can be lawful access. Nonetheless, the report perhaps intends that lawful access will connote that there should not be any violation of technologically protection measures, such as, breaking of paywalls to access the data, or any other illegality in terms of piracy etc. It may also mean that the access to data must not violate the terms and conditions attached to copyrighted data, such as the terms of usage of a website (which may deny permission for AI training) or the terms associated with proprietary databases. However, which law will override the other in order to determine the legitimacy of access is unclear; the public law that allows AI training for all available data, or the private law created by contractual terms that may explicitly deny any such use.[41] If AI developers have already scooped the data, which was denied for ingestion to AI training in the first place before the implementation of the proposal, it is unclear if it can still be called lawful. From the compliance perspective, the model may force the developers to subscribe to paywalled content, instead of relying on pirated materials, increasing the cost of production (which the DPIIT proposal is chiefly intending to avoid in the initial phases for start-ups and small players). Additionally, policy maker may note that most of the data scraping that is will be considered lawful access, as envisioned by the DPIIT hybrid model, and only time will decide about it, when the model is implemented and subjected to judicial scrutiny. In any case, the policymakers should attempt to remove the ambiguity on the aspect of lawful access by defining and characterising the instances of both lawful and unlawful access, bearing in mind the development of data scraping technology and its current capabilities.

F. The Resurrection of Retroactive Application

The suggestion to implement the proposed framework retrospectively would perhaps, add another layer of controversy. The report notes that many AI developers would have already ingested copyrighted content for AI training, and therefore it is only fair to share the appropriate revenue post commercialisation with the rightsholders. It narrates how many AI companies are already successful and generating huge revenues. Therefore, on the grounds of fairness and accountability, AI developers may be expected to pay the prescribed royalties for prior usage, as a corrective measure. Such an application of law, the report suggests, will only create a level playing field for all the players, new and old. While all this may seem good on paper, there would be many difficulties in its implementation: *First*, models which have discarded the dataset after training will be at a loss of any trail. *Second*, the extent of past training will be very complex to calculate, besides

[41] (There may be occasions in which contractual terms of a scrapped website had restricted the scarping and despite of the same, the AI developer had obtained the data on the pretext of fair use – a claim not supported at a later stage. This may create issues of legitimacy of the data obtained. This is an unsettled issue in many jurisdictions. In this context, the public policy to allow data for training (with compensation, of course) will complicate the situation. If it will be considered lawful is an aspect that the policy makers must answer to bring clarity).

being objectionable from the perspective of developers, who have hitherto acted under the assumption of fair use, citing the absence of any Indian court rulings to the contrary. *Third*, it may be argued that retroactive application of the law would be arbitrary and burdensome for existing players. The argument of vested commercial expectations may be brought up, as it was not explicitly forbidden to use protected data for training.[42] Moreover, historically, the Indian judiciary, through its pronouncements, has not favoured the imposition of laws and policies for retrospective imposition of civil liabilities,[43] and therefore, the proposed model's retroactive application can be struck down on the grounds of being unreasonable and expropriating. Therefore, a careful reconsideration of bringing retrospective application of the payment should be made, in case the AI players are finding it enormously difficult to comply.

IV. CONCLUSION

Law cannot stop the development of technology; rather, it catches up with the technology and readjusts when needed. It is undeniable that generative AI and its applications are growing exponentially, regardless of the support it receives from domestic law and policy. While it is laudable that the DPIIT working paper seems to be committed to providing a conducive environment to adopt AI in India and benefit our society through its transforming potential, the means and methods of achieving the same must be critically examined in order to ensure that its theoretical foundations remain coherent and normatively sustainable. The hybrid model proposed by DPIIT has no precedent in any other jurisdiction and is perhaps very novel in its approach. However, it has picked up the discrete ingredients from the domain of Indian copyright law, like the role of copyright society (existing for a few decades), statutory licensing that exists in radio broadcasting, and rate-setting methods and has coalesced them to provide a unique model, with a new institution to support it. The stated goal of balancing AI innovation and copyright cannot be sustained on a weak doctrinal foundation. Therefore, it is imperative that all possible threats to such a model's theoretical basis must be carefully discussed before its implementation. Nonetheless, the proposal will give the necessary direction and momentum to bridge the regulatory gap that is created by the emergence of the generative AI technology and democratize the policy making in this domain.

---

[42] *Retrospective Amendments and the Doctrine of Vested Rights: A Judicial Perspective*, TAXTMI (Aug. 13, 2024), https://www.taxtmi.com/tmi_notes?id=1284.

[43] S. Krishnan, *Retrospective Application of an Amendment*, TAXMANN, https://www.taxmann.com/research/income-tax/top-story/105010000000016563/retrospective-application-of-an-amendment-experts-opinion.

The most beneficial aspect of the DPIIT paper, and its proposed model is the explicit recognition of the rightsholder's interest in the AI ecosystem, which was hitherto neglected by the legislative attempts in the other jurisdiction. In the ever-expanding notion of fair use, a blanket TDM exception (like that of Japan) would have further curtailed the negotiating power of the rightsholders, giving a free ride to the technology companies over the use of the copyrighted data. Such a system is neither desirable nor sustainable in the case of copy-reliant technologies like TDM. The EU model of TDM, with an option to opt out, is also problematic, due to the advanced capabilities of modern TDM technologies that can easily manipulate the technical barriers and restrictions. Moreover, given the difficulties of compliance in the Indian scenario, enforcing opt-out may only remain on paper and will never be fully realised. Accordingly, it is a positive step that the content creators are being treated as a protagonist in the scheme of things and the AI companies are nudged to respect their data, and its associated IP rights, while developing products around them. Increased cost of compliance cannot be an excuse to violate IP rights, as, in any case, the cost would devolve to the end users of AI, and society at large, which may have the proclivity to absorb the cost. On the contrary, misappropriating the copyrighted content for training AI and selling the trained model would be akin to unjust enrichment,[44] as the content holders will be disincentivized with respect to their existing creations, apart from being discouraged to bring further creative expressions – something the society should never afford. While it is true that the AI and copyright saga will not necessarily create a zero-sum game (where one's loss gets exactly compensated by other's gain), a flawed system that encourages free-riding will also help no one; neither the interest of society, nor the sustainability of the AI systems in the long run. Therefore, perhaps diverging from the blanket TDM mechanism is the right choice for the Indian scenario at this point. Nonetheless, the proposed model may bring up many challenges as discussed, and therefore, some of its elements should be reconsidered before making the final blueprint of amending the copyright law.

Though international treaty compliance may not pose much of a problem for the proposed model, except as a potential ground of criticism, providing an adequate redressal mechanism to benefit rightsholders in case of disputes will correct the deviation from the Berne three-step philosophy and keep it sufficiently treaty-compliant. *First,* it should be ensured that the bargaining power of the rightsholders is never diminished amidst the influential AI developers and their business prerogatives. *Second*, the issues in pricing must be appropriately tackled, considering the market

[44] Yangzi Li & Jyh-An Lee, *Unjust Enrichment as a Remedy for AI's Unauthorised Use of Protected Data*, COMMON L. WORLD REV. 2 (2026).

price of AI training data and regularly factoring in their variations with time in determining the royalty rate. *Third*, the price fixing must be further democratised, ensuring adequate representation from all the stakeholders. Thirdly, the operational difficulties in the distribution of the revenue to the rightsholders must be progressively removed, though some hiccup will invariably occur at the beginning. *Finally*, the retroactive application of the payment mechanism can be reconsidered if the compliance burden becomes debilitating for a growing number of AI players. Only when these suggestions are embedded can the proposed mandatory licensing model become acceptable and successful, by protecting society's interests and reasonably settling the tussle between rightsholders and AI developers. Only then can the object of harnessing the transforming potential of the AI technology be truly realised.