O.P. JINDAL GLOBAL
INSTITUTION OF EMINENCE DEEMED TO BE
UNIVERSITY
A Private University Promoting Public Service

CYRIL SHROFF
CENTRE FOR
AI, LAW & REGULATION

Jindal Global Law School
India's First Global Law School

# COMMENTS

*on*

# WORKING PAPER ON GENERATIVE AI AND COPYRIGHT - PART 1

## TITLED 'ONE NATION ONE LICENSE ONE PAYMENT: BALANCING AI INNOVATION AND COPYRIGHT'

# A. Table of Contents



# B. List of Abbreviations & Definitions

| ABBREVIATION/ TERM | DEFINITION |
|---|---|
| **AI** | Artificial Intelligence as defined under Article 3(1) of European Union's Artificial Intelligence Act, i.e. "a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, inferring from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that influence physical or virtual environments" |
| **AI Developers** | Persons involved in, or responsible for, the creation and development of generative AI tools/systems |
| **AI Systems** | Generative AI tools, i.e. systems which can create new output in the form of text, images, sound, videos, etc. based on user prompts |
| **Act** | The Copyright Act, 1957 |
| **CMOs** | Collective Management Organisations |
| **Committee** | The committee appointed by the Department for Promotion of Industry and Internal Trade (DPIIT) to examine the intersection of generative AI and copyright. |
| **CRCAT** | Copyright Royalties Collective for AI Training |
| **EU** | European Union |
| **TDM** | Text and data mining |
| **TPM** | Technological Protection Measures |
| **UK** | United Kingdom |
| **US** | United States of America |
| **Working Paper** | Working Paper on Generative AI and Copyright - Part 1 titled 'One Nation One License One Payment: Balancing Ai Innovation And Copyright' |

# C. About the Cyril Shroff Centre for AI, Law & Regulation

The Cyril Shroff Centre for AI, Law & Regulation (the "CSCAILR") is a leading hub and centre of excellence that strives to produce high-quality research and actionable policy recommendations. We engage directly with policymakers and connect academic scholarship with policy implementation with the objective of enabling India to develop state-of-the-art frameworks for the regulation and governance of AI.

CSCAILR is a pioneering initiative at Jindal Global Law School (JGLS), O.P. Jindal Global University (JGU), India's top-ranked and globally reputed law school. Built under a generous endowment by Mr. Cyril Shroff, one of India's leading lawyers, CSCAILR is uniquely positioned to partner with academics, policy makers, think-tanks, law firms, the judiciary and businesses, to produce insights tempered with multi-stakeholder viewpoints.

# D. Executive Summary of Comments

Upon carefully reviewing and deliberating upon the evidently well-meaning proposal forwarded by the Committee through the Working Paper, we note that while the proposal does seem to theoretically balance the seemingly divergent interests of AI Developers, rightsholders and the public at large, there are certain shortcomings and lacunae. In particular, it appears that the Committee's proposal, if implemented, would lead to creation of several organisations and bodies, incurrence of high administrative costs and effort, and financial and administrative burdens on AI Developers. However, despite the cost, we believe it will not proportionately benefit human creators of copyrighted works; which is the primary objective of the proposal. On the other hand, imposition of payment and disclosure obligations may position India as an unfavourable jurisdiction for AI training, with AI Developers easily side-stepping obligations by ensuring their backend operations, including training, takes place outside the territory of India. Even if massive amounts of royalties are obtained from AI Developers, given that it will likely be divided among millions of creators, the final amount to reach rightsholders would be insignificant, thus, not fulfilling the Committee's expectation of incentivizing human-creation of works. Undoubtedly, there is a need to protect human creators of works, however, the Working Paper's proposed manner of achieving that objective may not be practicable or appropriate. In this respect, the following issues should be considered:

(1) **AI Training & Copyright Infringement:** The Working Paper's proposals seem to proceed under the assumption that the training of AI Systems on copyright protected works amounts to infringement. The retroactive, mandatory, blanket statutory license is proposed as a remedy to this issue. However, no Indian court has yet ruled on the issue of infringement during AI training, and AI Developers have an arguable defence of non-infringement and fair dealing under Section 52 of the Act. This has been discussed in our detailed comments below. Proceeding to lay down an elaborate proposal for mandatory statutory licensing without first establishing infringement may not be appropriate as any person who is not infringing copyright cannot be forced to comply with requirements for a statutory license.

(2) **Applicability & Territoriality:** The Copyright Act, 1957 applies only within the territory of India. Therefore, an act of infringement must take place, wholly or in part, within the territory of India for the Act to apply. During the AI training stage, which is distinct from the AI deployment and output generation stage, the only plausible claim for infringement may lie against the reproduction, storage and adaptation of works as databases are copied, stored and fed into an algorithm which converts the works into machine-readable language, tokenises the data, extracts patterns and/or assigns weights. These are back-end operations that take place within the servers and computer systems of the AI Developers. Therefore, as long as these servers and computer systems dealing with the training are located outside India, no actionable infringement under the Act can be established, and the statutory licensing mandate proposed by the Working Paper cannot apply.

Given that no other major jurisdiction currently has a similar or more onerous obligation relative to what has been proposed in the Working Paper, it is likely that AI Developers would shift their backend operations and infrastructure out of India, or would refrain from setting up within India, just to avoid the proposed revenue sharing obligation. This would, however, not prevent these AI Developers from accessing the Indian consumer market for AI products and services later. Thus, the Working Paper's proposal may inadvertently have a negative economic impact. It may also impede India's goal of developing indigenous AI Systems.

**(3) Lawful Access as a Pre-Requisite:** The Working Paper's proposal of providing a retroactive, blanket statutory license for AI training but imposing the lawful access requirement only prospectively is fundamentally unfair as it would give a clean chit to AI Developers who trained their systems on pirated or unlawfully accessed works as long as they comply with the proposed revenue sharing and data disclosure obligations for the retroactive statutory license. While the Working Paper states in passing that the issue of lawful access in the past can be left to judicial decisions, this is not compatible with the retroactive nature of the license. If the license is retroactive with no positive obligation of lawful access, courts would be bound by the amended legislation to find non-infringement.

Another inadvertent pitfall of the lawful access requirement could be the reduction of publicly accessible works as more rightsholders could put their works behind paywalls, technological protection measures and restrictive licenses to ensure that their works cannot be freely used for AI training purposes. This would be to the detriment of the public at large who would otherwise benefit from open access to the works.

**(4) Retroactive Application of the Statutory Licenses & the Obligation to Pay Royalties:** The proposed retroactivity of the amendment to the Copyright Act envisaged by the Working Paper is likely to face serious legal challenges from both rightsholders and AI Developers. For those who believe training of AI Systems on works amounts to infringement, it is a loss of potential injunctive relief and damages, while for those who believe training would not infringe copyright, it is an added obligation for actions done before the law changed. Both AI Developers and rightsholders may contend that they could have conducted themselves differently if the proposed law was known to them at the relevant time, and the shifting of legal requirements retroactively in such a scenario may be considered innately unfair.

**(5) Rate Setting Mechanism:** The proposed constitution of the Rate Setting Committee does not have sufficient representation from the rightsholders and the AI Developers, i.e. the two stakeholder groups who would be most affected by the proposed rate setting. This government-administered committee, and the proposed review of rates every 3 years, may not be optimal, especially in a field as dynamic and fast-transforming as AI.

**(6) Rate Setting Model:** The imposition of payment obligations on all AI Developers generating revenue without any minimum revenue threshold may act as a burdensome obligation on India's lagging AI industry. Generation of revenue itself does not mean profits, especially in the high-investment AI industry. Moreover, imposition of payment obligations based on global revenue instead of domestic revenue would be against territoriality and potentially disproportionate, especially since revenue generated has no rational nexus with the harm caused to rightsholders.

Imposing a royalty on the global revenue may disincentivize international players from setting up AI training infrastructure, servers and operations within India, leading to a negative impact on jobs and the economy. Moreover, it may also disincentivize AI Developers from training their AI Systems on Indian data, leading to underrepresentation and potential bias.

**(7) Disclosure Obligations & Apportionment of Royalties between Categories of Works:** The Working Paper is not clear about the width and depth of data to be disclosed by AI Developers, the manner of detecting non-compliance and the repercussions for false or incomplete disclosures. Any disclosure requirement must have a reasonable nexus with, and be proportionate to, the purpose to be served. Naturally, not all that is present in datasets which are used for AI training are protected by copyright. To require AI Developers to identify and segregate copyright protected works from other data would be impracticable and arduous, if at all possible.

**(8) Works Database for AI Training Royalties & Royalty Distribution:** The royalties collected by the CRCAT and distributed among its member organisations are to be further distributed among persons who have registered their works with such organisations. However, given that this would include members as well as non-members, and the sheer volume of copyrighted works that could potentially be covered, we believe that it would be impracticable, if not impossible, for member organisations to verify authorship, assess copyrightability, avoid duplication, keep a track of registrants, and also devise methods to fairly distribute the royalties.

If royalties are distributed on a pro-rata basis, the sheer number of registrants is likely to ensure that the in-hand royalty per registrant would be too negligible to incentivize human creativity or provide sustenance. If royalties are distributed based on assessed values of the work, the process of determining value would be extremely challenging, highly subjective, potentially biased, and likely unfair.

# E. Detailed Comments

Recent developments in AI technology have brought about a paradigm change in various aspects of life and has impacted every industry. The creative industry is no exception. Laws must evolve with changing times, and so, there is an urgent need for Indian Copyright Law to adapt to the new realities of how works are created, protected and utilized. However, while we adapt the Copyright Act, 1957 to the AI-age, and attempt to balance important countervailing interests and objectives, we must keep in mind that we do not create legal stipulations and mechanisms which are impractical, misaligned with objectives, or have unintended negative consequences.

We appreciate the Committee's thoughtful and well-researched efforts aimed at balancing the objective of providing equitable, low cost and practicable access to large quantities of data to AI Developers on one hand, and the protection of the countervailing interests of rightsholders on the other. It is true that we must strive to protect and foster our rich cultural and creative landscape by preserving human creativity: one of the avowed functions of copyright law is to do just that. However, any solution to the copyright issues raised by generative AI must not only be fair but must also be legally sound, practicable and efficient.

We understand that the Working Paper proposes to grant a retroactive, mandatory, blanket statutory license to all copyright-protected works that have been used, or will be used, for training of AI Systems. This is coupled with an obligation upon the AI Developers, who have commercialised AI Systems, to pay a percentage of their annual revenue as royalty to a centralized, non-profit, collecting body, proposed to be named 'Copyright Royalties Collective for AI Training' ("CRCAT"). Thereafter, the royalties collected by CRCAT is to be distributed among member organisations, i.e. copyright societies and not-for-profit collective management organisations ("CMOs"), each of which is responsible for any one class of works. These organisations are tasked with creating policies for, and implementing, the distribution of royalties among registrants, i.e. members and non-members who have registered their works with the organisation.

Keeping this in mind, we forward the following comments on the Working Paper which detail our understanding of the issues that may arise with the framework proposed by the Committee—

## 1. AI TRAINING & COPYRIGHT INFRINGEMENT

1.1. Traditionally, licensing works for training AI Systems has not been considered as one of the several ways in which rightsholders can exploit their works. However, the same could have been said for most new forms of exploitation at their inception. The fact that a particular manner of exploitation was not envisaged at the time of enactment of the Copyright Act, 1957, or even when a particular work was created, should not disentitle rightsholders. The virility of law lies in its robust adaptation to changed circumstances. Under the Act, any reproduction or adaptation of a work is the sole right of the holder of copyright[1], and therefore, unless an individual's act of reproduction or adaptation of the work falls within the limited exceptions provided in the Act, such as fair dealing[2], it would amount to infringement[3]. Notably, even the right to electronically store, or authorise the storage of, a work vests exclusively with the copyright holder[4].

1.2. According to Section 52(1)(a), an act is not infringement if it amounts to a "*fair dealing with any work, not being a computer programme, for the purposes of— (i) private or personal use, including research…*"[5]. AI training usually involves reproduction of copyrighted works after scraping, the storage of the reproduced copies in servers, possible transmission of such copies, and the adaptation of these copies as they are algorithmically converted to machine readable language and, ultimately, reduced to abstractions devoid of any expressive fixation through tokenization and assigning of weights. The act of reproducing, storing and adapting the works as aforesaid should be considered infringement, if they are not covered under Section 52(1)(a), since the other exceptions contained in Section 52 are very unlikely to apply.

1.3. The terms 'private use' and 'personal use' have not been defined under the Act, but the inclusion of 'research' within the scope of Section 52(1)(a) suggests that the end-goal of the private or personal use need not be non-commercial. Indian courts have held that commercial exploitation of works simpliciter does not preclude a fair dealing defence[6]. In case of transformative use, even the reproduction of the whole work may be excused under Section 52[7]. Any individual who reproduces a textbook for personal or private use may do so with the objective of gaining knowledge or skills which they may ultimately put into practice in their business, profession or job and thereby gain commercial value. This has never been outside fair dealing under Section 52(1)(a).

1.4. If we look at training of AI Systems, the act of training itself could be considered private or personal as there is no communication to the public of the reproduced or adapted copy at that stage. We concede that if AI Systems later

1.5. Now, if the scraping, reproducing, storing, and adaptation of works for AI training can be considered fair dealing protected under Section 52(1)(a), any amendment to the Copyright Act, 1957 which mandates payment for use of works for training of AI Systems must either be limited to activities which would not be considered fair dealing, or go as far as creating an obligation to pay notwithstanding any successful fair dealing defence. The Working Paper, insofar as it assumes the existence of infringement by AI Developers, may require reassessment.

1.6. Another aspect to be considered is that AI Developers, often, do not themselves scrape the internet, but instead purchase datasets compiled by third parties[8]. In this case, the sellers of the datasets reproduce, store and transmit the works for commercial gain to the AI Developers. It is unlikely that such activity would be considered as fair dealing under Section 52(1)(a), as commercial exploitation of works are not exempt from infringement simply because the end-use is private or personal[9].

1.7. This may create a duality where AI Developers who access the works, store them, adapt them and use them for training may have a fair dealing defence, while those who purchase datasets from others may not. In a case where the former are exempt from making royalty payments under the statutory license proposed by the Working Paper and the latter are not, the most likely outcome is vertical integration of businesses to ensure that the same entity scrapes data, stores it, tokenises it, and ultimately trains AI Systems on it. This would also cause disadvantage to smaller businesses lacking the capital or the resources to integrate all activities, thus, skewing the law in favour of larger, more resourceful businesses.

## 2. APPLICABILITY & TERRITORIALITY

2.1. The Copyright Act, 1957 applies only within the territory of India[10]. While Section 62 provides a long-arm jurisdiction to courts, since the Act itself is limited to India, infringement must take place within India for the Act to apply[11]. While it is true that the legality of an imported copy is determined by the law of the importing country[12], this would not affect AI Developers who reproduce and/or adapt works during training outside the territory of India, i.e. whose servers responsible for scraping, reproducing and storing copies of the works, or computer systems, responsible to adapt the works to machine-readable forms or converting the works to abstractions, exist outside India.

2.2. The fact that the trained AI model is later deployed in India would not constitute infringement at the input stage, though a material reproduction of the works in the output could still be considered infringement. However, infringement in the output generated by AI Systems is a different matter not covered under the Working Paper, and is, therefore, outside the scope of these comments.

2.3. Thus, the obligation to pay under the statutory licenses proposed by the Working Paper would arise only upon those who reproduce, store, or transmit works within India without a voluntary license, and not upon those who carry out AI training outside India and only supply their services to India after training. Thus, unless all major jurisdictions follow suit and adopt a mechanism more onerous, or at least as onerous as what has been proposed in the Working Paper, most AI Developers would prefer to move, or keep, their operations of training their AI systems outside Indian territory. This would negatively impact India's economy and data sovereignty goals. While already lagging in the AI race, with not a single indigenous large language model comparable to the several models developed in the U.S. and China, this would make India seen as an unfavourable jurisdiction for AI training by the developers.

## 3. LAWFUL ACCESS AS A PRE-REQUISITE

(a) Pitfalls of giving prospective effect to the lawful access requirement:

3.1. There can be no disagreement with the Working Paper's position that lawful access to works must be a pre-requisite for their use in training AI Systems[13]. However, giving prospective effect to the lawful access requirement while giving retroactive effect to the statutory licenses[14] would create an anomaly. This could result in AI Developers taking advantage of the proposed statutory licenses despite not having lawfully accessed the works used for training at a time before the proposed amendment to the law. This would be unfair to the rightsholders who would have otherwise had a much stronger case of infringement in the absence of the proposed retroactive statutory licensing.

3.2. While giving retroactive statutory licenses to AI Developers, as proposed by the Working Paper, is likely to face stiff resistance in courts from both, rightsholders and AI Developers (as discussed later), which may not be advisable, if the Parliament does proceed to give retroactive effect to the licensing provisions, they should give retroactive effect to the pre-requisite of lawful access as well.

(b)Reduction in free public access:

3.3.Another issue which must be considered is the possibility that rightsholders begin closing off their works behind paywalls, registration requirements, restrictive licensing agreements and the likes to prevent lawful access for training of AI Systems, which could, as an inadvertent outcome, reduce the accessibility of these works for the public as well, thereby proving counterproductive. An opt-out mechanism for rightsholders permitting them to prevent use of their works for AI training may solve this issue, even though it may reduce the amount of data available for training AI Systems. However, an opt-out mechanism built into the statutory licensing system would still be less detrimental to data availability for AI training as compared to an opt-out mechanism for a text and data mining exception (as existing in the EU) as the incentive of royalty payments would likely lead fewer rightsholders to opt out.

## 4.RETROACTIVE APPLICATION OF THE STATUTORY LICENSES & THE OBLIGATION TO PAY ROYALTIES

4.1. The Working Paper suggests that the obligation to pay royalty, and consequently the statutory license, should be given retroactive effect.[15] This is likely to be resisted by both rightsholders and AI Developers as—

- From the rightsholders' perspective, it would take away, or at least substantially limit, their right to sue, or continue their lawsuits, for infringement and potentially obtain injunctive relief and/or damages. Rightsholders may also argue that they would have not permitted lawful access to their works if they knew that this would result in a statutory license for AI training in the future through a retroactive legal amendment; and

- From the AI Developer's perspective, it would retroactively impose significant obligations which did not exist at the time when actions leading to the impositions and training were taken without clarity on whether their actions were in violation of law at the time. There are grounds to claim non-infringement and/or fair dealing which AI Developers might want to press before courts. AI Developers may submit that they would have conducted themselves differently if the retroactive law was known to them earlier in time. For instance, they could have kept servers and their activities dealing with works outside the territory of India and escape payment obligations.

4.2.What is proposed by the Working Paper is not merely a clarificatory amendment being given retrospective effect, but a substantial change in the law brought about retroactively. This would likely leave pending lawsuits relating to this matter partly, if not wholly, infructuous, and would not permit jurisprudence on pertinent issues raised by the lawsuits to develop. Academic interest aside, this would give a clean slate to infringement, if indeed what AI Developers have been doing categorises as infringement, even though the activities took place before the law was amended, possibly with knowledge and wilful assumption of the legal risk that came with it. An unprecedented retroactive amendment of this nature could potentially open a Pandora's Box, sending the message that legal violations may be absolved by the legislature if enough people do it.

## 5. RATE SETTING MECHANISM

5.1. The proposed constitution of the Rate Setting Committee is skewed strongly in favour of government nominees, with just one member each representing AI Developers and the rightsholders.[16] Thus, the people most impacted by the rate setting are underrepresented and will have little say on the rates, making this, essentially, a government-determined rate. Such massive government interference and abandonment of market forces would be antithetical to India's economic outlook of liberalisation and free markets. We recommend that the Rate Setting Committee should have equal number of representatives for the rightsholders and for the AI Developers, with one representative from the government to facilitate agreement, or in the absence, act as a tiebreaker.

5.2.While the Working Paper suggests that the rates would be susceptible to judicial review, it is unclear what the standard of judicial review would be. While a writ court may assess the rates for arbitrariness under Article 14 of the Constitution, there is always the possibility of different, parallel actions before various high courts resulting in conflicting decisions. We would recommend that a statutory amendment in line with the proposals of the Working Paper must also provide for judicial review by the Supreme Court of India so that litigants need not resort to writ jurisdiction under the Constitution and decisions bind the entire country equally. The grounds of challenge, including arbitrariness, discrimination, etc. should be statutorily specified. Nevertheless, the rate setting mechanism is likely to be a litigious issue. The above proposal would at least limit it to one court.

5.3.The proposed review of rates every 3 years does not appreciate the fast pace of technological changes, where paradigm transformations are taking place in months not years. The Rate Setting Committee must be alive to the challenges posed by ever-changing commercial realities and should act as a standing committee which reviews the rate continually. The Rate Setting Committee should be empowered to modify rates as and when it deems necessary.

## 6. RATE SETTING MODEL

(a) Absence of minimum revenue thresholds

1.1. We appreciate that arriving at a rate setting model for use of works in AI training is a complex and novel task. Nevertheless, it must be kept in mind that: (i) Revenue does not equal profit, and the fact that an AI system has been commercialised may not mean that it has begun, or indeed will ever begin, generating profits. Sharing of a percentage of revenue in such a case may have a disproportionately higher detrimental impact on cash-strapped or small-scale startups and small businesses, while also increasing the compliance burdens. This can act as an additional barrier to entry for smaller businesses who are already faced with a severe competitive disadvantage relative to businesses with more data, more capital and more experience in the industry. We would second MeitY's suggestion of at least having a minimum revenue threshold, to be notified by the Central Government and revised periodically, for the revenue sharing obligation to kick-in.

(b) Use of global revenue as the measure

1.2. We would also submit that the use of global revenue as the basis of determining the amount to be paid to CRCAT may be a disproportionate measure, especially since this would disregard:

(i) Territoriality: Fixing an amount as a percentage of global revenue disregards that all or substantial portions of training of AI systems may have taken place in other jurisdictions. Copyright law is territorial. Therefore, it would be unfair to require businesses to pay a portion of their global revenue to the CRCAT even if some portion of the training, and therefore some reproduction of copyrighted works, took place in India. Limiting this obligation to Indian revenue would at least establish some nexus between commercial benefits reaped in India and use of works for training in India.

(ii) Proportionality: Charging a proportion of global revenue as royalty under the proposed hybrid statutory licensing scheme would disproportionately affect AI Developers who do not earn much revenue from the Indian market, or have only a fringe presence in the Indian market, with primary operations being carried out in other countries, and who have not substantially trained their AI Systems in India.

(iii) Impact on AI Investment & Development: Charging a royalty based on global revenue may discourage international players from setting up servers or carrying out training, research and development of AI Systems in India, which would lead to lesser job creation, lesser foreign investments and general detriment to the economy.

(iv) Underrepresentation of Indian Data & Bias: The disincentive created by charging royalty on global revenue may lead businesses to refrain from training their AI Systems on Indian data, which could eventually enhance the chance of bias and exclusion of India and its works in the output.

## 7. DISCLOSURE OBLIGATION OF AI DEVELOPERS & APPORTIONMENT OF ROYALTY BETWEEN CATEGORIES OF WORKS

7.1. The Working Paper proposes imposing a statutory disclosure obligation upon AI Developers (called the "AI Training Data Disclosure Form"), which would be submitted to CRCAT and would contain details as prescribed by the Central Government through rules.[17] However, the Working Paper does not clarify the depth of information to be shared or the repercussions of non-compliance, and the reference to recital 107 of the EU AI Act exacerbates this.

7.2. If disclosure is mandated, the law would also need to set out the repercussions of non-disclosure or incorrect disclosure. Would AI Developers who fail to comply lose the statutory license proposed by the Working Paper, and thus, be liable for copyright infringement? Would such persons still be able to avail a fair dealing defence under Section 52? Would courts be empowered to direct discovery in lawsuits filed against AI Developers to check the accuracy of their declarations, or would independent audits into declarations be mandated? All of these are questions which will naturally arise and must be given thoughtful consideration.

7.3. Any disclosure requirement must have a reasonable nexus with, and be proportionate to, the purpose to be served. Naturally, not all that is present in datasets which are used for AI training are protected by copyright. To require AI Developers to identify and segregate copyright protected works from other data would be impracticable and arduous, if at all possible. We understand that one of the purposes for the disclosure obligations is to ascertain what proportion of each category of work has been utilised for training an AI System, so that CRCAT can proportionately divide the royalty collected among organisations representing the rightsholders of different categories of works. However, this glosses over the following issues:

- Datasets will contain works which are not protected by copyright, either because they are not the subject matter of copyright protection or protection over which has expired.

- Datasets may contain several copies of the same work, and there can be duplication of works across multiple datasets used in training. Accounting for that would be arduous.

- Works often have overlapping copyrights, especially in case of works which are built using underlying works (for instance, a cinematograph would have underlying literary works, such as, movie scripts, dialogues, lyrics of songs, as well as sound recordings, artistic works, dramatic work, photographs, each of which are separate categories of works under the Act).

7.4. Therefore, disclosure of categories and quantities of works used would not fairly represent how much of each category of copyright-protected work has been used in the training process, and therefore, the remuneration distributed based on this would naturally be non-representative and inaccurate.

## 8. WORKS DATABASE FOR AI TRAINING ROYALTIES & ROYALTY DISTRIBUTION

8.1. The Working Paper envisages that the members of CRCAT will be able to maintain a database of the various works of Registrants, including non-members.[18] Given the low threshold to obtain copyright protection over a work in India[19], there are billions of copyright protected works existing, and millions created every day. Photographs clicked, texts and emails written, musical notes, sound recordings, and the likes are easily produced in the hundreds and thousands each day. Technically, much of that is protected by copyright and the author can legitimately seek to have them registered with the CRCAT member for the category of work.

8.2. With no meaningful qualitative barrier, and no requirement to show that a work was actually used by AI Developers, the scheme proposed by the Working Paper[20] is, essentially, a free-for-all. This will be further exacerbated by the easy and free access to generative AI systems, which allow people to create text (even literary texts in the form of poetry, prose, stories, etc.) and even photographs, sound recordings and videos with little effort. While preventing duplication of works or detecting fake works through technological measures, such as, blockchains, digital watermarking, content fingerprinting and the likes, as has been suggested in the Working Paper, it assumes not only the existence of technological knowledge and access with creators, but also the existence of generally acceptable technical solutions for authentication. To expect that any organisation will be able to effectively verify and manage applications from a large number of people hoping to get on the bandwagon is unrealistic.

8.3. Notably, the Working Paper cites a report commissioned by the U.S. Copyright Office to show the capabilities of collective rights management organizations (CMOs).[21] However, the same report gives numerous reasons for unclaimed royalties, and explains the difficulties faced by CMOs in identifying usage, matching used works with owners, and distributing royalties.[22] Costly processing of information, cost of identifying and paying owners being disproportionate to the payable royalty amount, lack of sufficient resources, and multiple conflicting ownership claims are some of the prevalent reasons forwarded. The report acknowledges significant royalty leakages and practical difficulties faced by CMOs in distributing royalties. Notably, the said report relates to CMOs dealing only with rights related to music recordings, and the CMOs surveyed are all membership-based organizations, i.e. the CMOs are only responsible for their members. The Working Paper goes a step further and requires CRCAT members to cater to non-members as well, as long as they register their works with the CMOs. Thus, the task facing such CRCAT members would be exponentially more challenging.

8.4. There is no technology or management system existing today which can perform the task of weeding out AI-created content, detect duplication, and verify the provenance of each work registered with the CRCAT members, at least sine a disproportionately high cost. Even if such systems existed, the sheer number of likely registrants will ensure that any royalty distributed among the registrants will become insignificant. The stated objective of incentivizing human creativity will likely not be achieved.

8.5. As an alternative to pro rata distribution, the Working Paper does suggest apportionment of royalty among registrants based on assessed value of the works.[23] However, this is also impracticable given the subjective nature of works and the sheer volume of works likely to be registered. A straitjacket formula, no matter how insightful, is unlikely to be fair or fool proof. The use of parameters such as website traffic, licensing instances, citations, awards and social engagement illustratively suggested in the Working Paper can also not fairly give meaningful insights as to value for all types of works even within the same broad category.

8.6. Moreover, the value attributed to a work by society may not have a linear co-relation with the value the work may have for AI training. Therefore, adjudging value of the work and apportioning royalty based on indicators which show value attributed by society but do not represent value to the AI training process may not be fair.

8.6. Moreover, the value attributed to a work by society may not have a linear co-relation with the value the work may have for AI training. Therefore, adjudging value of the work and apportioning royalty based on indicators which show value attributed by society but do not represent value to the AI training process may not be fair.

8.7. Finally, ensuring a fair and efficient working of CMOs and other CRCAT members will be a challenge. Some of the maladies that have plagued collective licensing organisations, such as mismanagement, inefficient functioning, corruption, and lack of transparency[24] are likely to affect CRCAT members as well. The governance of these bodies, if left to the larger members, would skew royalty distribution mechanisms in favour of such members, especially if royalty distribution is based on value assessment of the works. The Working Paper's proposal leaves significant leeway to membership-based CMOs to determine the royalty distribution method, not only for its members but also non-member registrants, who will have no representation or say in the governance of such organisations. Whether these organisations would work for the benefit of non-member registrants will be suspect.

# F. Conclusions

Ultimately, the primary reason for the Working Paper's proposal of a hybrid statutory licensing model for AI training is to give monetary incentives to human authors of works and ensure human-created works continue to be produced. At least that appears to be the primary reason for rejecting text and data mining exceptions akin to Japan, Singapore, the U.K. and the European Union. To meet this objective, the Working Paper proposes a complex model which will not only require significant legal amendments, which are likely to be challenged in courts (especially if made retroactive), but will also require the creation of several new organisations and complex mechanisms. This will require substantial effort, time and capital, and even thereafter, the working of these organisations and mechanisms may still pose new challenges. On the other hand, the proposal will impose significant financial liabilities on AI Developers who would be obligated to set aside a portion of their revenue for this purpose, acting akin to a tax or rent on their business operations. We have already spoken of how such a law, if implemented only in India, could further push AI Developers to move (or keep) their backend and training operations outside Indian territory, but continue to access the Indian market of AI product consumers s, thereby affecting jobs and hurting the Indian economy.

However, despite the potential pitfalls and challenges, the proposed system does not guarantee meaningful incentives. Revenue collected by the CRCAT from AI Developers, no matter how large, would likely be distributed among too many people to be meaningful incentives in their hands. Large publishers, production companies and other aggregators may be able to obtain relatively sizeable financial benefits, but this is unlikely to percolate to the actual individuals behind the creation of the works. Even such amounts are unlikely to be large enough to incentivize large players and aggregators to employ or remunerate human creators, especially if AI-generated works prove to be good substitutes. In short, we find it unlikely that human-created works will continue to be produced simply incentivized by royalties received for training of AI Systems on such works, nor do we believe that human-created works will stop being produced in case no royalties are received for AI training. The real challenge lies in the output generated by AI Systems, and the extent to which the law, consumer preferences and market forces will permit AI-generated works to substitute human-created works. Therefore, we believe the focus of law and policy should be concentrated toward the output side of AI, while the development of AI Systems should be permitted unimpeded by additional financial burdens at this time, especially if such burdens could potentially make India an unfavourable jurisdiction for AI development.

# References


[1] *See*, Section 14, The Copyright Act, 1957.
[2] Section 52, The Copyright Act, 1957.
[3] Section 51, The Copyright Act, 1957.
[4] *See*, Sections 14(a)(i), 14(c)(i)(A), 14(d)(i)(B) & 14(e)(i), The Copyright Act, 1957.
[5] Section 52(1)(a), The Copyright Act, 1957.
[6] *Super Cassettes Industries Ltd. v. Hamar Television Nework Pvt. Ltd*, 2010 SCC OnLine Del 2086 at para 8.
[7] *The Chancellor Masters and Scholars of the University of Oxford v. Narendera Publishing House*, (2008) 38 PTC 385.
[8] OECD (2025), *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, OECD Artificial Intelligence Papers No. 33, at p. 19. Available at https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf.
[9] *Rupendra Kashyap v. Jiwan Publishing House*, 1996 SCC OnLine Del 466 [holding at para 21— "*The words 'research or private study' have been substituted by the words 'private use including research' by the Copyright (Amendment) Act, 1994 (Act 381 of 1994). What is contemplated is a defense to the person conducting research or private study who while doing so, if dealing fairly with a literary work, may not incur wrath of the copyright having been infringed. But, if a publisher publishes a book for commercial exploitation and in doing so infringes a Copyright, the defense under section 52(1)(a)(i) would not be available to such a publisher though the book published by him may be used or be meant for use in research or private study. The defense raised by defendants 1 and 2 based on Section 52(1)(a)(i) is not available to them and the plea so raised has to be rejected.*"]
[10] Section 1(2), The Copyright Act, 1957.
[11] *Blueberry Books v. Google India (P) Ltd*, 2013 SCC OnLine Del 6585 [holding at para 30—*"The Court is of the view that while Section 62(2)CA is a long-arm provision, the CA itself operates only through the territory of India. For the infringing activity to constitute a cause of action, it must be shown to have taken place in the territory of India.")*
[12] *See*, *Penguin Books Ltd. v. M/s. India Book Distributors*, 1984 SCC OnLine Del 190.
[13] Section 5.2, Working Paper at p. 62.
[14] *Ibid.*
[15] Section 5.4.3, Working Paper at p. 66.
[16] Section 5.4.1, Working Paper at p. 64.
[17] Section 5.5.1, Working Paper at p. 67.
[18] Section 5.5.3, Working Paper at p. 69.
[19] *See*, *Eastern Book Co. v. D. B. Modak*, AIR 2008 SC 809.
[20] See, Section 5.5.6, Working Paper at p. 74.
[21] Susan P. Butler, 'Collective Rights Management Practices Around The World - A Survey of CMO Practices to Reduce the Occurrence of Unclaimed Royalties in Musical Works', April 2020, Butler Business & Media LLC, available at https://www.copyright.gov/policy/unclaimed-royalties/cmo-full-report.pdf.
[22] *Id.* at pp. 10-14.
[23] Section 5.5.5, Working Paper at p. 71-72.
[24] *See*, Jonathan Band & Brandon Butler, 'Some Cautionary Tales about Collective Licensing', (2013) 21 (3) Michigan State International Law Review 687 at p. 690-709.


O.P. JINDAL GLOBAL
INSTITUTION OF EMINENCE DEEMED TO BE
UNIVERSITY
*A Private University Promoting Public Service*

CYRIL SHROFF
CENTRE FOR
AI, LAW & REGULATION

Jindal Global Law School
*India's First Global Law School*

O.P. Jindal Global University, Sonipat Narela Road,
Sonipat-131001, (Delhi NCR), Haryana, India

jindalglobaluni jindalglobaluni jindalglobaluni jindalglobaluni jguvideo