SFLC.IN
K-9, Second Floor, Birbal Road,
Jangpura Extension,
New Delhi-110014
(tel): +91-11-43587126
www.sflc.in

Comments on DPIIT Working Paper on Generative AI and Copyright (Part 1)

Executive Summary

On December 9, the Department for Promotion of Industry and Internal Trade released a working paper on Generative AI and Copyright- One Nation One License One Payment Balancing AI Innovation and Copyright ("**Working Paper**"). The Working Paper proposes a mandatory blanket licensing model that would allow AI developers to lawfully access all publicly available data for AI training. Copyright owners/creators will have a statutory remuneration right and will receive compensation in the form of royalty through a centralized authority and collective management societies.

SFLC.in brief concerns are along the following lines:

- In the mandatory blanket licensing model, creators lose autonomy over their content and are forced to take part in AI training without an opt-out or opt-in option. There are also no safeguards for personality rights.
- The mandatory blanket license model ignores Free and Open Source Software (FOSS) and open licenses, which are built on "voluntary sharing" and "share-alike" principles. Mandatory licensing allows proprietary AI systems to absorb open-source work without respecting downstream license obligations. It also forces creators and developers who intentionally share their work for public benefit to accept compensation and undermine FOSS ideologies.
- The proposed flat, revenue-linked remuneration model does not offer fair or meaningful compensation to creators. Creators are asked to surrender their works upfront while compensation remains speculative, depending on the AI company's revenue and are likely negligible.
- It narrows the scope of existing copyright infringement remedies (damages, injunctions) by limiting it to unlawful access and royalty disputes.
- The Working Paper proposes solutions for AI training without defining what constitutes copyright infringement in generative AI models.

SFLC.IN
K-9, Second Floor, Birbal Road,
Jangpura Extension,
New Delhi-110014
(tel): +91-11-43587126
www.sflc.in

SFLC proposes the following solutions:

- Adopt an opt-in license model for AI training, so that creators have autonomy to choose whether their content is available for AI training, and respect the existence of different licensing models such as open-source licenses.
- Introduce upfront payment for training data, plus revenue-based royalties, once the AI company generates revenue. This prevents creators from bearing the risk of AI experimentation.
- Limit the mandatory blanket license model to public-interest sectors like education, healthcare, and non-commercial research, and it must be narrowly tailored, purpose-bound, and justified by demonstrable public interest.

SFLC.IN
K-9, Second Floor, Birbal Road,
Jangpura Extension,
New Delhi-110014
(tel): +91-11-43587126
www.sflc.in

Introduction

The Working Paper proposes unique mechanisms to address copyright issues arising from the use of copyrighted content in AI training and examine whether India's copyright regime adequately addresses the issues, and provide recommendations. The Committee recommends a "Hybrid Model", built around a mandatory blanket license for AI training combined with a statutory right to remuneration for copyright holders. By introducing a mandatory blanket license with a statutory right to remuneration, the Committee is trying to move away from consent-based and case-by-case licensing, towards a system that guarantees access while ensuring creators are paid. The Hybrid model also aims to solve the lawful data access problem of AI developers, by introducing blanket licensing. This would mean all content that is publicly available online, without bypassing technological protection (freely accessible without paywalls, logins, or other access restrictions), is accessible for AI training. The statutory remuneration plays out through a centralized collecting entity known as the 'Copyright Royalties Collective for AI Training ("CRACT"),' with its governing body consisting of balanced representation from each member organization ("Copyright Society"/ "Collective Management Organizations"/ "CMO"). The entity is proposed to collect royalties from AI Developers based on rates fixed by a government appointed committee, on behalf of its member organizations and the unrepresented sectors which do not have CMOs are represented by government-nominated representatives at the Governing Board of CRCAT, and distribute these royalties to its member organizations (i.e. CRCAT members).

There are different schools of thought on whether AI training itself constitutes infringement or whether infringement should be assessed only at the output stage when substantial similarity arises. Nuanced discussion of what constitutes copyright in AI use, its input and output, and how it works in this regard, is essential before prescribing a licensing model. There are several complexities in deciphering the work that is used for an AI output, with its non-measurable status makes matters complex and calling for a mandatory licensing model is a simplistic solution. All other considerations flow from this argument- the rationale for not having an opt out mechanism, and assisting the AI developers by reducing transaction costs, compliance burden, etc. Even though Part 2 of the Working Paper will be in the context of copyright and AI, some background discussion, especially around the unavoidable question of what makes for copyright infringement/ copyrightable, is essential to discuss what can be a part of CRCAT's royalty collection business. This is crucial as certain works can be treated as part of copyright exception without its express mention, given the scope of Section 52 of the The Copyright Act 1957. The Working Paper proceeds on the assumption that training requires regulatory intervention without

addressing if training results in infringement. Before redesigning copyright law itself, the Working Paper should be more clear on what issues it is actually trying to solve and why this shift is necessary and how will the shift align with the core objectives of copyright law which is to incentivise creativity by granting exclusive economic and moral rights to the creator.[1] The Working Paper definitely takes away the exclusive rights of creators and it also raises concerns on whether creators will be fairly incentivized.

Key Concerns on the Features of the Working Paper are as follows:

1. Mandatory Blanket License

i) Lack of autonomy for creators: The mandatory blanket licensing framework proposed under the hybrid model fundamentally alters the nature of copyright by converting it from an exclusive property right into a statutory license with uncertain remuneration. Copyright is designed to provide creators with authorial autonomy and control over the use of their works, including the right to refuse permission. But the mandatory blanket license model replaces consent with compulsory participation, the framework treats creators paternalistically i.e. by assuming that creators are incapable of making informed decisions about their own works and that the State must intervene on their behalf. It also does not provide the option of neither opt-in nor opt-out to AI training to the creators. Thus, creators are denied the option to make informed choices based on moral, reputational, or economic considerations. The lack of opt-in/opt-out option also means that creators may permanently lose control over their content once it is used in AI training.[2] What is presented as "licensing"; therefore seems to be legitimizing appropriation without an opportunity to negotiate the terms of the license.

ii) Excessive scope of lawful access: Under the mandatory blanket license model, any access to publicly available data on the internet without technical protection is within limits for the AI developer to use for AI training and such access is "lawful access". It is unclear whether contractual obligations, platform terms of service or existing license requirements will continue to apply. The model also risks creators who currently have not deployed technical protection measures and paywalls to lock up their content to protect their work. This will ultimately reduce public access to creative content and undermine the open internet. Small creators may not have the means to resort to technical protection and may remain exposed in comparison to large publishers who can shield their archives easily and may even charge hefty

[1] Comments by Dr. Tania Sebastin, Assistant Professor of Law, School of Law, VIT.
[2] Higa H, Bedikian S and Costa L, "The Right to Be Forgotten Is Dead: Data Lives Forever in AI" *Tech Policy Press* (May 20, 2025) https://www.techpolicy.press/the-right-to-be-forgotten-is-dead-data-lives-forever-in-ai/

amounts for subscription. Over time, this may reduce the diversity and availability of creative works online as large publishers will control gated ecosystems. Moreover, lawful access also does not clearly address cached content, scraped user-generated material, platform-hosted works, or content accessed under platform terms that prohibit automated reuse. The lawful access feature under the Hybrid Model doesn't seem to be balanced, it clearly provides AI developers leeway in accessing content while shifting the burden of protection onto creators.

iii) Lack of compatibility with Berne Convention and TRIPS Agreement: Article 9(2) of the Berne Convention, Article 13 of TRIPS and Article 10 of the WCT establishes a three-step test for any copyright limitation or exception. Any such exception must: (i) apply only in certain special cases; (ii) not conflict with the normal exploitation of the work; and (iii) not unreasonably prejudice the legitimate interests of the right holder.[3]

- The Working Paper proposes an economy-wide mandatory blanket license applicable across all sectors, works, and categories of creators for commercial AI training. This cannot reasonably be characterised as a "certain special case" within the meaning of the Berne three-step test.[4] The Working Paper does not explain how such a broad, cross-sectoral licensing mandate satisfies this threshold.[5]
- Under the three-step test, any exception must not conflict with the normal exploitation of the work. Under copyright and tort principles, authors are entitled to claim both direct damages and loss-of-profit damages. AI training directly interferes with these interests by enabling the rapid creation of competing outputs and derivative works. Further, AI-generated content trained without creators' consent may directly compete with original works, thereby undermining their ordinary commercial exploitation.
- Third, exceptions must not unreasonably prejudice the legitimate interests of authors. The TRIPS Panel in 2000[6] clarified that "prejudice to the legitimate interests of right

[3] Calderon R, "Ai Training through Copyrighted Works as Infringement: Perspectives under the Berne Three-Step Test and the Pane Fair Use Test and Plausible Solutions Rubé N Calderón" (*Franklin Pierce School of Law*) https://law.unh.edu/sites/default/files/media/2024-09/6-ruben-essay_.pdf
[4] Rosati E, "Copyright Exceptions and Fair Use Defences for AI Training Done for 'Research' and 'Learning,' or the Inescapable Licensing Horizon" 16 European Journal of Risk Regulation 961 https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/copyright-exceptions-and-fair-use-defences-for-ai-training-done-for-research-and-learning-or-the-inescapable-licensing-horizon/752DF1DB564AD1EDFE23BA8BB1110802
[5] Knights R, "Limitations and Exceptions under the 'Three-Step Test' and in National Legislation – Differences between the Analog and Digital Environments" (WIPO) http://www.wipo.int/edocs/mdocs/copyright/en/wipo_da_mvd_00/wipo_da_mvd_00_4.doc
[6] "WTO" (*dispute settlement - the disputes - DS160*) https://www.wto.org/english/tratop_e/dispu_e/cases_e/ds160_e.htm

holders reaches an unreasonable level if an exception causes, or has the potential to cause, an unreasonable loss of income to the right holder." In the absence of opt-in or opt-out mechanisms, the mandatory licensing model risks unreasonably prejudicing creators' legitimate interests. While the Working Paper proposes statutory remuneration, it remains doubtful whether meaningful income can realistically be distributed to all creators. The cumulative effect may discourage creative production over time and drive creators away from publishing their content or from creation itself.

Although societal benefits of AI may be invoked, such benefits cannot override the requirement that creators retain the ability to derive fair economic returns from their works.

iv) Personality Rights : Generative AI tools can, on simple prompts, create images, videos, and voice outputs that closely resemble an existing person's persona. Such imitation, apart from being used to generate fake or misleading content, has also paved the way for deepfakes, AI-morphed images, and other forms of digital exploitation. Many celebrities have filed lawsuits to limit such unauthorized use of their features including image, likeness, voice and other distinctive traits.[7] Under Section 38A and Section 38B of the The Copyright Act, performers are granted exclusive rights and moral rights allowing them to control how their performances are reproduced and to object to any distortion or misuse. In *Arijit Singh v. Codible Ventures LLP*[8], the Bombay High Court granted an ex-parte ad-interim injunction in favour of singer, Arijit Singh. The popular singer filed a suit against AI companies for the unauthorised commercial use of his personality through AI tools. The Court recognised that Arijit Singh is a well-known singer with goodwill and celebrity status in India, and that his personality traits are entitled to legal protection against third-party exploitation. The Court held that AI tools which allow users to convert any voice into that of a celebrity without consent violate the celebrity's personality rights. A person's voice is a core part of their identity and public image, and unauthorised use or manipulation of it, amounts to an infringement of their right to control their likeness. Such use also enables misleading and commercial misuse of a celebrity's identity. The Court further observed that creating and commercially using AI-generated audio or video content in the plaintiff's name, voice, image, or likeness without consent could seriously harm the plaintiff's career and livelihood. Allowing such tools to continue without any restriction would not only cause economic harm but also create risks of misuse for deceptive or harmful purposes.

[7] Chopra P, Sony RAL and Chopra S, "Generative AI, Copyright and Personality Rights: A Comparative Legal Perspective" (2025) 6 Legal Issues in the Digital Age 23 https://lida.hse.ru/article/view/28690
[8] Arijit Singh *v*. Codible Ventures LLP**,** 2024 SCC OnLine Bom 2445.

SFLC.IN
K-9, Second Floor, Birbal Road,
Jangpura Extension,
New Delhi-110014
(tel): +91-11-43587126
www.sflc.in

The Working Paper thus overlooks the growing risk of Generative AI applications infringing personality rights, particularly those of celebrities, performers, journalists, and public figures. It does not acknowledge the increasing misuse of AI tools to replicate voices, faces, mannerisms, and identities, nor does it propose any framework to prevent or remedy such violations.[9] For data that involves a person's identity such as their voice, face, or likeness, compensation alone is not enough and people must have the right to give informed consent before their personal attributes are used to train AI systems. This protection should apply even when such data is publicly available online.[10] Without these safeguards, the proposed framework risks normalising the large-scale use of human identity for AI development without meaningful control or accountability. At the same time, any safeguards must also preserve space for satire, parody, caricature, criticism, and legitimate artistic expression, so that protections against AI impersonation do not unintentionally restrict free speech or creative commentary. The Committee may also refer to Denmark's Copyright Act Amendment, which proposes to grant individuals rights over their own likelinesses.[11]

v) Implications for FOSS and Open Licenses: The Working Paper disregards existing licensing systems of the Free and Open Source Software (FOSS) and open knowledge communities. licenses like GNU GPL are built on voluntary sharing, downstream conditions, and community norms rather than monetisation. These licenses grant users the freedom to run software for any purpose, study it, modify it, and redistribute it in original or modified form.[12] Most open-source licenses, including GNU GPL and Creative Commons licenses, are built on the principle of downstream reciprocity i.e. any derivative or redistributed work must be shared under the same license terms and attributed to the original creator. This "share-alike" structure is central to open source ecosystems. The mandatory blanket license model disrupts such existing license frameworks. When an AI developer trains proprietary models using open-source or openly licensed material, the resulting models and outputs are not distributed

[9] Anturkar A, "Deepfakes And Dignity: The New Battle For Celebrity Rights In India" *Live Law* (November 27, 2025) https://www.livelaw.in/articles/celebrity-rights-personality-rights-india-deepfake-misuse-legal-framework-article-21-311287

[10] "AI and Personality Rights: Legal Implications" (International Journal of Law Management & Humanities) ttps://ijlmh.com/paper/ai-and-personality-rights-legal-implications/

[11] Fleck G, "Denmark Leads EU Push to Copyright Faces in Fight Against Deepfakes" *Tech Policy Press* (October 7, 2025) https://www.techpolicy.press/denmark-leads-eu-push-to-copyright-faces-in-fight-against-deepfakes

[12] Wickramarachchi V, "Understanding the Common Licensing Models in Open Source Projects" (*Bits and Pieces*, August 29, 2021) https://blog.bitsrc.io/understanding-the-common-licensing-models-in-open-source-projects-e07375979c16

under the original license terms. This bypasses the terms of the open source licenses and the creators intent for downstream use. As a result, value moves only in one direction: from community-built resources to closed commercial systems. This weakens copyleft protections and affects the collaborative model open-source depends on. The Hybrid Model thus allows open work to be absorbed into proprietary systems without respecting downstream licensing obligations. There is also an ongoing litigation in the US, where the plaintiffs have challenged GitHub for breach of contract for violation of open-source license for AI training among other claims.[13] The Working Paper has not considered such nuances.

The Working Paper also raises a fundamental ideological dilemma for developers who create and share their work for public benefit. Such creators, under the Hybrid Model are forced to accept remuneration simply because their code exists and is technically accessible to AI systems, in the absence of an opt-out option. The Hybrid Model mandatory blanket licensing imposes a one-size-fits-all framework that treats all creative and technical labour as monetisable input for AI training, it doesn't respect diverse licensing choices and creator intent to permit free use. Thus, it undermines the philosophical foundations of the FOSS community and disregards the rich history of voluntary collaboration of open-source ecosystems.

Moreover, unlike in the case of J. Doe 1, et al. v. GitHub[14] in the US, it is unclear whether a license holder can approach the court for an infringement claim, because under the Working Paper the AI developer has accessed the AI training in a "lawful manner".

vi) Impact on Journalism and News Media: The mandatory blanket licensing framework raises concerns for journalists and news media organisations, journalism and news media already operates at a structural disadvantage under intellectual property law. Much of the labour-intensive work involved in journalism, such as field reporting, verification, and investigation is not fully protected by copyright, even when it costs significant financial and human resources. At the same time, journalism revenues have been steadily declining due to platform dominance and the rise of generative AI systems that aggregate, rewrite, and reproduce news content without adding meaningful analytical value. The Hybrid Model further weakens the position of journalists by removing the right to withhold content from AI training. This compulsory inclusion legitimises extraction of journalistic work in the name of innovation. Without stronger safeguards, such as the right to refuse AI training, transparency

[13] J. Doe 1, et al. v. GitHub, Inc., et al. Case 3:22-cv-06823 (N.D. Cal)
[14] idbi

obligations, and protections against reproduction, the proposed framework risks accelerating the decline of journalism rather than supporting it. The Working Paper also fails to recognise journalism as a public good. Any AI governance model must prioritise sustainable news ecosystems, editorial independence, and fair compensation, instead of treating journalistic content as raw material for AI systems.

vi) Lack of public interest: In the case of *Super Cassette Industries Ltd. vs. Music Broadcast*[15], the Supreme Court held that the compulsory licensing model is detrimental to copyright holders' right unless there is a demonstrable public interest for which such a license must be granted. The Court also emphasised that compulsory licensing model cannot be implied and mere financial gain arising from such model is only incidental. The Working Paper departs from this principle by moving to a mandatory blanket license which is compulsory in nature without identifying any demonstrable public interest for which the mandatory blanket license is required.

vii) Lack of meaningful remedies: Under current law, copyright holders enjoy exclusive rights to reproduce, distribute, perform, display, translate, and adapt their works, along with moral rights to claim authorship and prevent distortion, mutilation, or other prejudicial modification. Where infringement occurs, copyright owners retain full remedies, including injunctions, damages, and accounts of profits. The Working Paper disrupts this balance. By introducing mandatory blanket licensing without clearly defining what constitutes infringement in the context of AI, it effectively narrows the circumstances in which creators can seek enforcement. According to the White Paper, a copyright holder would only be able to claim infringement in limited situations such as unlawful access to training data (for example, bypassing paywalls or technical protections) or non-payment of royalties. This significantly weakens the broader spectrum of remedies currently available under copyright law and risks reducing enforceable rights to mere participation in a collective compensation mechanism.

2. Statutory Remuneration Rights and Royalty Distribution

We welcome DPIIT's recognition that creators deserve compensation when their works are used to train AI systems. Listed below are issues we see in the current remuneration model:

[15] Super Cassette Industries Ltd. vs. Music Broadcast (2012) 5 SCC 490

i) Lack of technical feasibility for determining attribution: In AI systems, learning happens through tiny fragments of data called "tokens." During training, billions of such tokens drawn from countless sources are blended together into model parameters. Once this process is complete, it is nearly impossible to identify which specific work is behind a particular output, or to measure how much any individual creator contributed to the model. Unlike music streaming or publishing, where each play or download can be tracked, AI outputs differ as they emerge from massive datasets. The Working Paper also assumes that AI Developers can meaningfully disclose how different "classes of works" were used in training through an *AI Training Data Disclosure Form*. At best, developers could provide long lists of training sources running into billions of entries.[16] But being named in such an undifferentiated list is not meaningful attribution. This problem becomes even more complex in real-world systems, where models are trained using other models, fine-tuned repeatedly, or improved based on lessons learned from earlier versions. As a result, royalty distribution under the proposed framework is a guesswork. A large portion of license revenue is likely to end up in unallocated "black box" pools because tracing rightsholders will be difficult or even impossible.[17]

ii) Disparity in compensations: The proposed flat revenue-share model also attempts to apply quantitative pricing (archive size, historical dominance, registration volume) to qualitative work by ignoring differences in efforts. This is likely to disadvantage small publishers, independent journalists, and emerging media platforms that lack large archives. Even optimistic projections show that global AI revenues, when divided among potentially billions of creators, would translate into only a few rupees or dollars per person annually. Such amounts cannot meaningfully compensate creators for the loss of control over their works or the market substitution caused by AI systems. There are risks of remuneration turning into a symbolic payment rather than genuine economic participation.

iii) Flat-revenue model is unfair to creators: The Working Paper proposes a flat percentage of global gross revenue as compensation for creators, with payment triggered only after commercialisation i.e. the AI developers do not have to pay for training if the Company does not generate revenue. A recent research study shows that around 95% of generative AI

[16] "Is There A Coherent Theory of Attributing AI Training Data?" (*Michael Weinberg*) https://michaelweinberg.org/blog/2024/07/15/ai-attribution-what-does-it-mean/
[17] "Reverse Robin Hood" (*Ex Machina*, December 17, 2025) https://exmachina.in/17/12/2025/reverse-robin-hood/

projects fail to deliver meaningful revenue impact. Most deployments stall or produce negligible returns.[18] This means creators are being asked to surrender their works upfront, without consent, while their compensation depends on a business outcome that, in most cases, never materialises. In effect, the model shifts all risk onto creators whose work is permanently absorbed into AI systems at the training stage, while AI companies pay only if and when they succeed commercially. This is not balance, this is structural inequality where creators finance experimentation by AI developers which will most likely result in no payment or negligible sums.

iv) Lack of Time Limits for Royalty: The Working Paper doesn't specify any clear timeframe for royalty obligations limits for the AI developers.

3. Copyright Remuneration Collective for AI Training (CRCAT) and Collective Management Societies (CMO)

India's previous experience with collecting societies have faced persistent allegations of opaque accounting, delayed or inconsistent royalty payments, institutional capture by intermediaries, and bureaucracy. Instead of empowering creators, these bodies often concentrate economic and administrative power in a small group of institutional actors and act as a deterrent to creators itself. [19] These governance failures were so serious that it prompted legislative intervention through the 2012 Copyright Amendment, which introduced mandatory re-registration, enhanced disclosure requirements, and periodic oversight. The Indian Copyright Office have also repeatedly intervened to deregister, restructure, or supervise collecting societies due to mismanagement.[20] Despite these reforms, systemic problems in royalty distribution and representation continue to persist.[21] History thus makes it clear that merely creating another

[18] Estrada S, "MIT Report: 95% of Generative AI Pilots at Companies Are Failing" *Fortune* (August 18, 2025) https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/

[19] Shweta Venkatesan, An In-Depth Assessment of the Intersection of AI and Copyright in India, November 2025, Esya Centre. https://static1.squarespace.com/static/5bcef7b429f2cc38df3862f5/t/69295eba637b0319a8125dae/1764318906141/Reupload+-+copyright.pdf

[20] Bal, Meghna. "Government Pushes for Greater Transparency in Proposed Changes to Copyright Rules." Financial Express, June 14, 2019. https://www.financialexpress.com/india-news/government-pushes-for-greater-transparency-in-proposed-changes-to-copyright-rules/1607684/

[21] Soni B, "Royalty Disputes in Music Streaming: Evaluating How Artists and Labels Navigate Copyright and Revenue Sharing" *Sonisvision* (March 31, 2025) https://www.sonisvision.in/blogs/royalty-disputes-in-music-streaming-evaluating-how-artists-and-labels-navigate-copyright-and-revenue-sharing

collective entity such as CRCAT or relying on CMOs do not guarantee fairness or efficiency but instead introduce an additional layer of intermediaries, whose incentives may not align with those of individual creators. Avoiding these pitfalls are critical to the success of CRCAT/CMO.

i) Unfavorable advantage for big creators: Large publishers and established rightsholders are more likely to register works and navigate CRCAT systems, while vast segments of India's creative economy including regional journalists, small media platforms, artisans, textile communities, folk artists, educators, and informal creators may lack awareness or support required to participate. This risks their works being used for AI training even in the absence of representation at the CMO level or remuneration.

iii) Heavy Reliance on AI Developers Disclosure: The framework heavily relies on AI developers' self-disclosure of training data categories. There are no meaningful verification mechanisms suggested within the Working Paper. Considering the scale and opacity of model training pipelines, there can be possibilities of under-reporting or misclassification. Meanwhile, the valuation methodologies proposed such as assigning "high," "medium," or "low" value scores based on market indicators do not really bear any correlation to AI training.

iv) Lack of participation of creators: The proposed framework does not provide creators any meaningful role in determining royalty rates. Compensation levels will be fixed by a centrally appointed Rate Setting Committee, in which creators may be indirectly involved through copyright societies. Individual creators have no statutory right to participate in rate negotiations, approve valuation methodologies, or influence class-wise allocations. This top-down approach undermines creator autonomy and risks reproducing existing power imbalances, where large rightsholders dominate decision-making while independent creators remain excluded.

Suggestions

1. **Opt-In mechanism:** Given the concerns above for mandatory blanket licensing model, a consent based opt-in licensing model offers more balance between creators and AI developers. Copyright is, by design, an opt-in property right: creators decide when, how, and by whom their works may be used. Reversing this default through blanket access changes established copyright law itself. The opt-out regime as seen in the UK and Singapore and as discussed

within the Working Paper is unreliable as creators have to actively object (opt-out) to automatic access to AI training. This shifts the burden on the creator to be vigilant. By contrast in the opt-in model, a creator's work is treated as unavailable for AI training unless the creator explicitly consents.[22] Once a creator opts in, the work may be used for AI training and the creator becomes eligible for remuneration as prescribed in the Working Paper. Importantly, in the context of personality rights, an opt-in model ensures that individuals have a meaningful opportunity to consent before their voice, image, likeness, or other personal attributes are used in AI training.

Under an opt-in system:

- Creators can retain control over the use of their works for AI training.
- Different classes of creators (journalists, artists, software developers, performers, researchers) can choose licensing terms aligned with their values and economic realities.
- There is clarity on lawful access as the creator consents to the access.
- Allows different licensing communities to continue operating under their existing models. For example, Free and Open Source Software (FOSS) communities can continue using copyleft licenses.
- Licensing can be standardized through registries, collective mechanisms, or machine-readable permissions
- Given the technical difficulty of selectively removing data from trained AI models, any subsequent opt-out can only operate prospectively but content already ingested into models will continue to remain.[23]

Thus, this approach respects creator intent and aligns with international copyright norms. Innovation should not depend on forced participation. A consent-based opt-in model ensures that AI development proceeds responsibly, with creators treated as rightsholders and not as passive data sources.

2. **Upfront Payment for Training Data:** AI developers should be required to pay an upfront fee based on the size and nature of the business and the AI model , this must be considered as a basic cost of doing business, similar to infrastructure, compute, or cloud services. Creators'

[22] Katie Hookins, Opt-in vs Opt-out: Why Consent Frameworks for AI Training Data Matter TrueRights, (Noovember 25, 2025)
https://truerights.com/knowledge-hub/opt-in-vs-opt-out-why-consent-frameworks-for-ai-training-data-matter
[23] Amanda Coelho Della, Fair Compensation for Copyrighted Data Used in AI training, Tilburg University (2024), https://arno.uvt.nl/show.cgi?fid=176944

works should not be treated as free raw material and should be compensated at the point of extraction, not only after commercial success. In addition to this upfront payment, a second layer of royalties should apply once an AI system becomes profitable. A dual model of payment will reduce risk for creators prior to being profitable (even when a model is not per se profitable, its presence in the market can still undermine creators' economic value by substituting their work) and will help creators from bearing the cost of potential AI development.

3. **Government to release training data:** The Government should proactively release publicly funded and government-held datasets for AI training under open and permissive licenses. Large volumes of valuable data already exist within public institutions including archives, educational materials, broadcast content, and publicly funded research but this remains inaccessible or locked behind administrative barriers. Making such datasets openly available would support AI developers and reduce dependence on privately owned copyrighted content.

4. **Sector Specific Exceptions:** If compulsory access to copyrighted content is mandated, it should be limited to public-interest sectors such as education, healthcare, and non-commercial research, where societal welfare is the government's primary responsibility. Any exceptional access must be narrowly tailored, purpose-bound, and grounded in demonstrable public interest and not just applied broadly to commercial AI development.

Conclusion

The Working Paper reverses the proper order of legal reasoning. Ordinarily, the Committee should first determine whether a legal right exists, whether that right has been infringed, and whether any statutory exceptions apply (which is expected to be addressed in Part 2 of the Working Paper). Only after answering these questions should legislative reform be considered to respond to the challenges posed by generative AI. The Committee's approach does not follow the said sequence and by departing from the sequence it has made it difficult to understand the depth of the problem that is being tried to solve and also to ascertain remedies to infringement.

Copyright laws emerged in a pre-AI era to address a specific economic imbalance: it was designed to protect creators' labour, skill, against the ease with which others could copy and distribute creative works. Its core purpose was to ensure that authors could recover the costs of creation in a world where reproduction was cheap and widespread. Today's large-scale machine

SFLC.IN
K-9, Second Floor, Birbal Road,
Jangpura Extension,
New Delhi-110014
(tel): +91-11-43587126
www.sflc.in

learning systems that ingest vast datasets, extract statistical patterns, and generate probabilistic outputs were never envisaged under this framework. Similarly, pre-AI licensing frameworks were also written to govern copying and dissemination, not model training, pattern extraction, or automated generation.

While copyright remains relevant for clear cases of infringement, it may not be sufficient to address the full range of issues raised by generative AI. Many of the challenges posed by AI such as labour displacement, market concentration, consent, transparency, and competition may fall outside copyright's traditional scope. Thus, new legal frameworks and licensing models may indeed be required for AI training. Hence, it is essential to first carefully assess a wider set of issues and harms emerging from AI and whether it falls within copyright's domain or whether dedicated AI governance mechanisms are required. Part 2 of the Working Paper will hopefully provide greater clarity on these foundational questions.

About SFLC.IN:

SFLC.IN is a donor supported legal services organization that brings together lawyers, policy analysts, students, and technologists to protect freedom in the digital world. SFLC.IN promotes innovation and open access to knowledge by helping developers make great Free and Open Source Software, protect privacy and civil liberties for citizens in the digital world by educating and providing free legal advice and help policy makers make informed and just decisions with the use and adoption of technology. SFLC.in has been granted Consultative Status with the Economic and Social Council of the United Nations (ECOSOC)