Journal of Law and Intellectual Property Rights
ISSN: 3049-0979, Volume 3 Number 1 (June' 2026) pp. 29–49
DOI: https://doi.org/10.66566/jlipr/2026.v3n1.04
Journal website: https://www.jlipr.in

Received: 15 Jan' 2026 | Revised: 12 Mar' 2026 | Accepted: 19 Mar' 2026 | Published: 01 Jun' 2026

OPEN ACCESS

Review Article

Balancing Innovation and Copyright Laws: A Comparative Analysis of AI Training Regulations and Roadmap for Legislative Future of India

Prithvika Survepalli* and Kanishk Goyal

Research Scholar, National Law Institute University, Bhopal, India.

Email: survepalliprithvika.ug@nliu.ac.in, kanishkgoyal.ug@nliu.ac.in*

ABSTRACT: ***The emergence of generative artificial intelligence has raised complex legal questions about the use of copyrighted content in training large language models (LLMs). While global attention often focuses on AI generated outputs, this paper shifts the lens to the training process itself, examining whether such use constitutes infringement under copyright law. Through a comparative study, the paper classifies jurisdictions into Pro Tech and Pro Law models. Pro Tech jurisdictions such as the United States, Japan, and Singapore permit text and data mining through statutory exceptions or expansive fair use doctrines. In contrast, Pro Law regimes, particularly within the European Union, prioritize licensing, transparency, and the protection of authorial rights. India, despite being a digital and AI hub, has yet to define its legal stance on AI training. The absence of clear statutory guidance creates a risk of unregulated appropriation of creative works, threatening both creators' rights and the integrity of its innovation ecosystem. This paper urges swift legal reform. Drawing from international models, it proposes a balanced Indian framework comprising amendments to the Copyright Act, licensing pathways, transparency obligations, and a dedicated AI and IP regulator. India must act now to shape a copyright regime that enables responsible AI development while safeguarding its rich creative economy.***

KEYWORDS: ***Artificial Intelligence, Deep Learning. Text-data Mining, Black-box, Copyright, Fair-use, Jurisdiction.***

INTRODUCTION

Law is the architecture of order, designed to govern the needs of society, but society is fluid, ever-evolving. With the rise of Artificial Intelligence (AI), we are witnessing not just a technological shift, but a redefinition of creativity, authorship, and ownership. AI is reshaping the field of Intellectual Property Rights (IPR), challenging foundational assumptions about 'who' or 'what' can create. Traditionally, legal scholars and courts have agreed that copyright protection is reserved for human creators, with technology seen merely as a tool, not a source of originality. This view reflects the core intent of copyright law: to reward human intellect.

However, this understanding is now being tested, as AI systems, especially those using deep learning and transformer architectures, begin to emulate aspects of human cognition.

While much debate has focused on whether AI-generated content should be eligible for copyright, a less examined but equally critical issue is the training stage of large language models (LLMs). Many of these models are trained on vast repositories of proprietary content, raising questions about consent, ownership, and the limits of fair use. Some countries permit such training under broad exceptions, while others remain cautious or lack clear policies altogether. In response, governments around the world are at alarm to craft legal frameworks to address this emerging challenge. India, standing at the intersection of a booming digital economy and a rich creative tradition, must take an approach that caters to its needs.

This paper examines the evolution of copyright law through landmark judicial decisions across jurisdictions, with a particular focus on how courts have addressed the legal challenges posed by AI training. The authors delve into the technical workings of the training process itself to evaluate whether requiring prior authorization for the use of copyrighted content is a just and proportionate demand. The paper also explores the diverging global approaches which have been broadly categorized as "pro-law" and "pro-tech" highlighting their motivations and practical outcomes. Drawing on this comparative analysis, the paper proposes a nuanced and forward-looking policy framework tailored to India's unique position at the intersection of innovation and intellectual property rights.

THE EVOLUTION OF COPYRIGHT IN A DIGITAL AGE

The evolution of copyright law is rooted in the effort to balance the rights of creators with the public interest in access to knowledge. The journey began with the Statute of Anne (1710)[1] in England, the first statute to grant authors, and not printers, the exclusive right to print their books for a limited term. This represented a significant shift from monopolistic control by the Stationers' Company toward a regime that recognized authorship as a legal and economic entitlement.[2] The concept of copyright was further conceptualized by the Berne convention[3] in which key principles were established; automatic protection upon creation, national treatment[4], and recognition of both moral rights (such as attribution and integrity)[5] and economic rights (like reproduction, adaptation, and public performance).[6] While the concept of *'originality'* became central to copyright eligibility, different jurisdictions laid down different standards for determining its threshold.[7] Early systems in the UK and other common law countries accepted the doctrine of *"sweat of the brow"*, where industrious labour alone could justify copyright.[8] However, this approach was eventually rejected by courts in favour

[1] The Statute of Anne, 1710 (U.K.).
[2] Alina Ng, Copyright Law and the Progress of Science and the Useful Arts (Edward Elgar 2011).
[3] Berne Convention for the Protection of Literary and Artistic Works, 9 September 1886, as amended on 28 September 1979.
[4] World Intellectual Property Organization, Summary of the Berne Convention for the Protection of Literary and Artistic Works (1886), available at https://www.wipo.int/treaties/en/ip/berne/summary_berne.html (Last visited on June 1, 2025).
[5] U.S. Copyright Office, Report on Moral Rights 27 (Washington, D.C., 1998).
[6] PACRA, "What are the economic rights of a copyright holder?" PACRA Service Information (), https://info.pacra.org.zm/what-are-the-economic- rights-of-a-copyright-holder/.
[7] Paul Goldstein, "Originality and Creativity in Copyright Law," 30 J. Copyright Soc'y U.S.A. 109 (1982), available at https://doi.org/10.2307/1191773 (Last visited on June 1, 2025).
[8] Hailshree Saksena, "Doctrine of 'Sweat of the Brow'," SSRN Electronic Journal (2009), available at https://ssrn.com/abstract=1398303 (Last visited on June 1, 2025).

of a creativity-based originality standard.[9] The US Supreme Court held that copyright requires '*independent creation*' coupled with '*modicum of creativity*'.[10] Another core principle is the *'idea-expression dichotomy'* which holds that copyright protects only the expression of ideas, not the ideas themselves.[11] Related to this is the *'merger doctrine',* which denies protection when an idea can only be expressed in a limited number of ways.[12]

The U.S. Act codified the fair use doctrine, which considers four factors: (i) the purpose and character of the use, (ii) the nature of the copyrighted work, (iii) the amount and substantiality used, and (iv) the effect on the market.[13] Fair use remains a contextual, case-by-case defense, allowing limited uses for education, commentary, news, and research.[14]

The court in case of *Burrow-Giles Lithographic Co. v. Sarony*,[15] affirmed that photographs are subject to copyright protection, recognizing the photographer's creative choices in composing the image. This decision expanded the scope of protectable works to include new forms of media. In *Anil Gupta v. Kunal Dasgupta*,[16] the court held that ideas can be copyrighted if a substantial amount of labour has gone into formulating it. In various other jurisdictions, providing ideas for the software creation does not constitute joint-authorship and that only expression in a tangible form is protectable. In the landmark case of *Naruto v. Slater,*[17] the court decided that Naruto, the monkey, could not own the copyright to the selfies because copyright law in the U.S. applies only to works created by humans. Since animals do not have the legal ability to be considered authors or hold copyrights, only human creators are eligible for such protections.

This decision has also been cited in various discussions of WIPO about the copyright status of works generated by artificial intelligence, highlighting the ongoing relevance of the human authorship requirement.[18] The Delhi High Court in a landmark case[19] rejected copyrightability of a computer-generated list without human intervention. In recent times, the court in the case of *Worhol v Goldsmith*[20] conceptualize 'fair use' and limit its threshold to focus upon the use of proprietary content without explicit permission or within the narrowed boundaries of authorised license. In *Thomson Reuters v. Ross Intelligence,*[21] the court found that Ross's use of Westlaw headnotes to train its AI did not qualify as fair use, emphasizing that the use was commercial, not clearly transformative, and posed potential market harm. Similarly, in *The*

[9] Robbin Singh, "Understanding the Concept of Originality under Copyright Law in India," Law Mantra Online Journal, Vol. 2, Issue 9 (2015), available at https://journal.lawmantra.co.in/wp-content/uploads/2015/08/11.pdf (Last visited on June 1, 2025).

[10] Feist Publications, Inc. v. Rural Telephone Service Co., 499 U.S. 340 (1991).

[11] Agreement on Trade-Related Aspects of Intellectual Property Rights, 1994, Art. 9.2, available at https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm (Last visited on June 1, 2025).

[12] Steven Ang, The Idea-Expression Dichotomy and Merger Doctrine in the Copyright Laws of the U.S. and the U.K., Vol. 2(2), inT'l J.l. & info. Tech., 111 (1994), available at https://doi.org/10.1093/ijlit/2.2.111 (Last visited on June 1, 2025).

[13] 17 U.S.C. § 107 (1976).

[14] Stanford University Libraries, "Fair Use Case Index," Stanford Copyright and Fair Use Center, available at https://fairuse.stanford.edu/overview/fair-use/cases/ (Last visited on June 1, 2025).

[15] Burrow-Giles Lithographic Company vs. Sarony, 111 U.S. 53 (1884).

[16] Academy of General Education, Manipal v. B. Malini Mallya, AIR 2002 Del 379.

[17] Naruto v. Slater, No. 16-15469 (9th Cir. 2018).

[18] WIPO, "Can the 'Monkey Selfie' Case Teach Us Anything about Copyright Law?" WIPO Magazine (2018), available at https://www.wipo.int/web/wipo-magazine/articles/can-the-monkey-selfie-case-teach-us-anything-about-copyright-law-40287 (Last visited on June 1, 2025).

[19] Navigators Logistics Ltd. v. Kashif Qureshi & Ors., CS(OS) 1599/2011, decided on 8 May 2018 (Del HC).

[20] Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 143 S. Ct. 1258 (2023).

[21] Thomson Reuters Enter. Centre GmbH v. Ross Intelligence Inc., 694 F. Supp. 3d 467 (D. Del. 2023).

New York Times v. OpenAI,[22] the court allowed key copyright infringement claims to proceed, pointing to instances where ChatGPT reproduced Times content and raising concerns about the impact on the Times' market, while dismissing some ancillary claims like unfair competition.[23]

Hence, the core tenets of copyright law, as consistently upheld by recent judicial interpretation, are rooted in safeguarding the creative autonomy and economic rights of authors. These principles are fundamentally at odds with the unauthorized use of proprietary content for AI training purposes. However, as technology evolves, it becomes essential to understand the underlying mechanics of AI training in order to make informed and balanced legal decisions that reflect the realities of this new paradigm.

THE MECHANICS OF MACHINE LEARNING: HOW AI SYSTEMS ARE TRAINED

Before we decide whether AI models should be trained on certain kinds of data, it's important to halt and understand how these systems actually learn. The term *training* might sound straightforward, but behind it lies a layered, data-intensive process where raw, often unstructured content is transformed into something that machines can use to simulate understanding. Only by unpacking this process can we start to have meaningful conversations about policy, ownership, and ethics.

The journey begins with Text and data mining (TDM) which is a process that's foundational to how modern AI systems, (especially LLMs), gather their learning material. TDM is essentially the large-scale collection and preprocessing of data from various sources.[24] These sources include books, academic articles, websites, social media posts, software code, and more. The idea is to expose the model to as wide a range of human expression and knowledge as possible. But contrary to popular belief, this data isn't stored as it is. It isn't memorized like pages in a notebook. Instead, it undergoes several layers of transformation before it becomes meaningful for machine learning.[25] When a model encounters raw text for the first time, it doesn't see paragraphs or ideas. It sees characters, symbols, and sequences. So, the first technical step is to process the data into a format the model can digest. This often involves *tokenization* that is breaking down language into smaller units like words, sub-words, or even individual characters.[26] For instance, the sentence *"The doctor prescribed a new medication"* might be broken into tokens like *["The", "doctor", "prescribed", "a", "new", "medication"]* or even sub-word units like *["The", "doc", "tor", "pre", "scribed"]* depending on the model's vocabulary structure. These tokens are then mapped into vectors—multi-dimensional numerical representations that capture their contextual usage across the training corpus.

What's fascinating here is that the model doesn't know what the word "doctor" means in a human sense. Instead, it learns that the token "doctor" often appears near words like "hospital," "nurse," or "patient." Over time, through exposure to millions of such co-occurrences, it begins

[22] The New York Times Company v. Microsoft Corporation, 1:23-cv-11195 (S.D.N.Y., filed Dec. 27, 2023).

[23] The Times's About-Face: NYT v. OpenAI, Harvard Law Review Blog (April 2024), available at https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/ (Last visited on June 1, 2025).

[24] D. V. Martens, F. Provost & D. Jensen, "Text and Data Mining: From Information to Knowledge" (2021) 15(1) Foundations and Trends® in Information Retrieval.

[25] R. Bommasani et al., On the Opportunities and Risks of Foundation Models (Stanford Center for Research on Foundation Models 2021), available at https://arxiv.org/abs/2108.07258 (Last visited on June 1, 2025).

[26] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021).

to associate certain tokens with specific contexts.[27] This is the essence of pattern recognition in language models. The problem with TDMs is that downloading works from the internet and storing them on servers even temporarily constitutes reproduction under U.S. copyright law, making such actions potentially infringing.[28] Additionally, the process of converting these works into tokens does not eliminate legal risk, as these tokens could theoretically be reversed into readable content, preserving the original expression in a different form. While these intermediate "invisible" copies are not accessed or read by humans, their legal status remains contentious, as their role in training still involves the unauthorized use of protected material.[29]

The deep learning architecture forms the foundational core of contemporary artificial intelligence systems. Transformers exhibit a high degree of proficiency in processing sequential data, rendering them particularly well-suited for complex natural language processing tasks.[30] Their strength lies in their ability to focus on relationships between words across a sentence or paragraph, regardless of their position.[31] For example, in the sentence *"Although the medication was effective, the patient experienced severe side effects,"* the model can weigh the relationship between "effective" and "side effects" despite the distance between them. It learns that language is full of nuance, contradiction, and layered meaning, not just simple cause-effect pairs.

Crucially, the model's learning doesn't involve memorizing exact examples. If a textbook on biology says *"Photosynthesis converts sunlight into chemical energy,"* the model doesn't store that exact sentence. Instead, it sees countless variants of that idea across different texts. It learns to statistically model the relationship between "photosynthesis," "sunlight," "conversion," and "energy." This is why, when prompted later, it can generate a paraphrased version like *"In plants, sunlight is harnessed to produce energy through photosynthesis."* The system doesn't recall; it reconstructs, based on probability.

This reconstruction ability is frequently misconstrued as genuine understanding; however, it is crucial to emphasize that large language models do not possess knowledge or comprehension in the human sense. Their strength lies in statistical generalization. Given a sufficient amount of diverse training data, the model becomes increasingly good at predicting what might come next in a sentence, how a question is usually answered, or how a specific idea is typically phrased.[32]

One way to think about this is through the analogy of language immersion. Imagine someone who doesn't speak French spending years reading French newspapers, listening to French radio, and watching French films. Eventually, they begin to internalize patterns of usage; common phrases, typical sentence structures, cultural references, even if they've never been formally taught. In a very loose sense, this is what happens during AI training. The model

[27] Emily M. Bender & Alexander Koller, "Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data" in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) (2020), available at https://aclanthology.org/2020.acl-main.463.pdf (Last visited on June 1, 2025).

[28] Yoshua Bengio et al., "A Neural Probabilistic Language Model" (2003) 3 Journal of Machine Learning Research 1137.

[29] Daniel Jurafsky & James H. Martin, Speech and Language Processing (3rd ed., Draft 2023), available at https://web.stanford.edu/~jurafsky/slp3/ (Last visited on June 1, 2025).

[30] Pamela Samuelson, "Generative AI and Copyright Law: An Overview" Columbia Journal of Law & the Arts (Forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4422946 (Last visited on June 1, 2025).

[31] IBM, "Transformer Model," IBM Think, available at https://www.ibm.com/think/topics/transformer-model (Last visited on June 1, 2025).

[32] Ibid.

becomes a statistical immersion learner, building fluency by exposure to massive, diverse inputs. That is the essence of how deep learning architecture aids this process.[33]

It's also worth noting that not all data is treated equally. This is because Pre-training usually involves a general dataset meant to teach broad language patterns. However, models can also be *fine-tuned* on narrower domains like legal documents, clinical notes, or programming manuals.[34] This is where things get more policy-relevant: the choice of training data, its source, licensing status, domain specificity and begins to shape what the model becomes good at, and potentially, where it starts to raise legal or ethical concerns.

The nerve of the issue is that this entire process gives rise to what is often called the "black box" problem.[35] As deep learning models grow in complexity, particularly when trained on vast and varied datasets, the internal pathways they use to arrive at outputs become increasingly opaque. This lack of interpretability poses serious challenges, especially in the domain of intellectual property rights.[36] If a model is trained on copyrighted material, and later generates content that closely resembles it, tracing how the original input influenced the output is nearly impossible.[37] Without transparency, it becomes difficult to determine whether the AI has merely generalized from patterns or effectively reproduced protected works. This blurring of lines not only complicates enforcement and rights attribution but also raises broader questions about accountability in machine-generated content.[38]

All of these complexities ranging from the legal ambiguity of text and data mining, to the transformation of content into tokens, to the inscrutability of deep learning models culminate in a tangled web of legal and ethical uncertainties around the training of AI systems. AI models learn by identifying statistical patterns across massive datasets, yet copyright law was built to govern human acts of direct copying not machine-driven generalization. This disconnect raises critical questions about what constitutes infringement, how "use" is defined, and whether tokenized representations and black-box outputs can (or should) trigger legal liability. This fundamental mismatch between technological capability and legal design has prompted varied responses across jurisdictions.

THE PRO-TECH PARADIGM: LEGAL FLEXIBILITIES FOR GENERATIVE AI

Every nation brings a unique regulatory philosophy to the table, from exceptions for machine learning to strict copyright enforcement, offering a comparative lens to explore how the world might balance innovation with rights protection in the age of intelligent machines. Certain jurisdictions have embraced a more permissive approach to AI training, prioritizing innovation and the advancement of machine learning over rigid adherence to traditional copyright norms. In contrast, others have been slower to adapt, maintaining stringent legal frameworks that may overlook the need for greater flexibility in light of rapidly evolving technological capabilities.

---

[33] J. Huang & J. Murphey, "Training AI on Copyrighted Data: Legal Risks and Technical Challenges" (2023) Harvard Journal of Law & Technology Digest.

[34] Daniel Gervais, "Copyright and the Data Dilemma" (2023) 46(3) Columbia Journal of Law & the Arts 383.

[35] Z. C. Lipton, The Mythos of Model Interpretability, arXiv preprint (2016), available at https://arxiv.org/abs/1606.03490 (Last visited on June 1, 2025).

[36] F. Doshi-Velez & B. Kim, Towards a Rigorous Science of Interpretable Machine Learning, arXiv preprint (2017), available at https://arxiv.org/abs/1702.08608 (Last visited on June 1, 2025).

[37] R. Binns, Algorithmic Accountability and Public Reason, Vol. 31(4), Philosophy & Technology, 543 (2018), available at https://doi.org/10.1007/s13347-017-0263-5 (Last visited on June 1, 2025).

[38] J. Burrell, How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms, Vol. 3(1), Big Data & Society (2016), available at https://doi.org/10.1177/2053951715622512 (Last visited on June 1, 2025).

As explained above, the process of training typically involves downloading and storing large volumes of proprietary data, from publicly accessible sources.[39] This content is then converted into numerical formats that the machine can process and learn from. Under conventional copyright frameworks, such as that of the United States, even this act of downloading and storing content for a non-transitory period is considered reproduction.[40] Furthermore, converting expressive works into numerical tokens can also be viewed as a form of copying, since these tokens could theoretically be reversed into something human-readable.

This raises a critical dilemma as these intermediate copies are not made for human consumption and, in most cases, are never seen or read by anyone. They serve only a functional role in enabling the AI model to recognize statistical patterns and linguistic relationships. Once trained, the model functions by predicting the next most likely token in a sequence, which over time results in the generation of new content. Importantly, models like GPT-4 are not designed to store or recall entire works.[41] In fact, the scale of the models is usually much smaller than the total volume of data they are exposed to during training. This design choice forces the model to learn through abstraction, rather than memorization.

The distinction between training data and the trained model is therefore fundamental. While there is a causal link between the two, the trained model is not a copy of the original material. It does not retain the full content of any particular work, nor do its outputs necessarily reproduce protected expression in a way that would satisfy the traditional test of substantial similarity in copyright law.[42] As a result, some legal systems have started to draw a line between expressive and non-expressive uses of copyrighted content, especially when the copied material is used solely for internal machine learning purposes and not for public display or commercial distribution in its original form.

Additionally, as TDM became central to both academic and commercial research, many jurisdictions began recognizing the need to clarify its status under copyright law. Legal scholars and policymakers in several countries began to view TDM as a legitimate non-expressive use deserving of exceptions or safe harbours. Traditional TDM methods focused on extracting factual or analytical insights, whereas generative AI can now produce digital content that resembles the form and format of its training data.[43] Even if these outputs do not meet the threshold of substantial similarity required to establish copyright infringement, they may still functionally compete with the original works or their authors, thereby creating potential issues for the fourth factor in the U.S. fair use test.[44]

Moreover, while many generative models are designed not to memorize specific works, studies have shown that some do reproduce content with noticeable resemblance to original training

[39] J. Burrell, How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms, Vol. 3(1), Big Data & Society (2016), available at https://doi.org/10.1177/2053951715622512 (Last visited on June 1, 2025).

[40] Copyright and Artificial Intelligence: Part 3 – Generative AI Training (Pre-Publication Version), U.S. Copyright Office (May 2025), available at https://www.copyright.gov/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf (Last visited on June 1, 2025).

[41] Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code, U.S. Copyright Office (Dec. 2024), available at https://www.copyright.gov/title17/title17.pdf (Last visited on June 1, 2025).

[42] John P. Hilton, Jr., The Future of the European Union's Unitary Patent: A Tale of Two Courts, Two Treaties, and Two Visions of European Integration, 54 IDEA: The Journal of the Franklin Pierce Center for Intellectual Property 269 (2014), available at https://ipmall.law.unh.edu/sites/default/files/hosted_resources/IDEA/p269.Hilton.pdf (Last visited on June 1, 2025).

[43] Andrea M. Matwyshyn & Brian D. Matwyshyn, AI Governance, Interoperability, and Interdependencies, 19 J. Intell. Prop. L. & Prac. 557 (2024), available at https://academic.oup.com/jiplp/article/19/7/557/7624901 (Last visited on June 1, 2025).

[44] Matthew Sag & Peter K. Yu, The Globalization of Copyright Exceptions for AI Training, 19(7) J. Intell. Prop. L. & Prac. 557 (2024), available at https://academic.oup.com/jiplp/article/19/7/557/7624901 (Last visited on June 1, 2025).

materials. This has led to real concerns among copyright holders and their advocates, who argue that the creation of hidden intermediate copies, even if used only for abstraction, should be subject to regulation. In response, they have urged legislative reform and legal challenges to restrict the use of copyrighted materials during AI training. These debates have shaped the evolution of copyright frameworks globally, especially in jurisdictions where innovation is treated as a policy priority.

Against this backdrop, certain jurisdictions such as, Japan, Singapore, and the United States have tended to respond with frameworks that either permit TDM explicitly or adopt flexible doctrines such as fair use to accommodate AI training. These countries focus less on the act of copying in isolation and more on the ultimate purpose and societal effect of such uses.

JURISDICTIONAL ANALYSIS

In assessing how different legal systems enable AI training while navigating the boundaries of copyright law, this section classifies pro-technology jurisdictions into three distinct categories. Each reflects a different legal strategy for accommodating non-expressive uses of copyrighted content in the context of generative artificial intelligence.[45]

The first group comprises jurisdictions that recognize broad judicial exceptions, most notably through doctrines of fair use or close functional equivalents. These systems rely primarily on case law to permit transformative or non-expressive uses of copyrighted works, including during AI training. Courts in such jurisdictions have tended to emphasize the purpose, nature, and effect of the use rather than focusing solely on the fact of technical reproduction. This judicially driven flexibility has allowed for innovation without requiring immediate statutory reform.

The second category consists of jurisdictions that have enacted explicit statutory exceptions for TDM or computational data analysis.[46] These jurisdictions recognize TDM as a legitimate and necessary activity for scientific research, digital innovation, and, machine learning. The existence of statutory exceptions provides clarity to AI developers by formally excluding certain training-related uses from the scope of infringement, often without requiring prior authorization from rights holders.[47]

The third category includes jurisdictions that are actively promoting AI development as a national strategic priority but have not yet updated their copyright frameworks to address the implications of AI training.[48] Despite the lack of express exceptions or fair use-style doctrines, these countries have adopted a de facto pro-technology stance by not aggressively enforcing copyright norms in this space. In some cases, this policy lag is intentional, reflecting a desire to attract AI investment or buy time for regulatory calibration. In other cases, it reflects the early stage of copyright reform debates in the national context. Each of these classifications represents a different legal pathway to supporting AI development, whether through courts, legislatures, or policy decisions.

a. Recognize broad judicial exceptions, most notably through doctrines of fair use or close functional equivalents

[45] Ibid.
[46] Ibid.
[47] Martin Senftleben, Compliance of National TDM Rules with International Copyright Law: An Overrated Non-issue? 53 IIC 1477 (2022), available at https://doi.org/10.1007/s40319-022-01266-8 (Last visited on June 1, 2025).
[48] Ibid.

The doctrine of fair use, previously introduced through the lens of the four-factor test,[49] takes on sharper relevance when viewed through the policy choices of jurisdictions like the United States and Israel, which have leaned toward a pro technology interpretation.[50] This framework, rooted in US copyright law[51] and echoed in Israeli jurisprudence,[52] has proven particularly adaptable to the demands of AI development. Its open-ended structure allows courts to assess whether AI related uses are transformative rather than simply derivative, making space for innovation without dismantling the rights of original creators.[53] What distinguishes these jurisdictions is not just the application of the four factors but how these are calibrated in the AI context. US courts have clarified that non expressive uses, such as the extraction of statistical associations from text, are inherently transformative and thus more likely to qualify for protection.[54] Israel's Ministry of Justice reinforced this interpretation by issuing an advisory opinion affirming that training AI systems on copyrighted data, without expressive reproduction, does not violate copyright norms.[55]

A key refinement here is the interpretation of the first and fourth factors. Both jurisdictions stress that even commercially motivated uses can qualify as fair if the end product does not substitute for the original or infringe on its expressive core. While discussing the fourth factor, i.e, the market impact, courts have been cautious. While they currently consider the training use non substitutive, they leave room for reevaluation should a clear licensing market for training data emerge.[56] These interpretations find resonance beyond US and Israeli borders. Countries like Singapore, South Korea, and the Philippines[57] have adopted similar flexible doctrines, if not identical fair use regimes, reflecting a growing international recognition that rigid copyright enforcement may stifle innovation in AI. Critics argue this flexibility introduces legal ambiguity, but supporters note it has enabled these jurisdictions to become global leaders in AI and data driven industries.[58]

[49] Folsom v. Marsh, 9 F. Cas. 342 (C.C.D. Mass. 1841.

[50] See Peter K. Yu, Fair Use and Its Global Paradigm Evolution, 2019 U.ILL. L. REV. 111, 128 [hereinafter Yu, Paradigm Evolution] ("Australia, Hong Kong, Ireland, Israel, Liberia, Malaysia, the Philippines, Singapore, South Korea, Sri Lanka, and Taiwan have already adopted or proposed to adopt the fair use regime or its close variants."); see also infra text accompanying note 117

[51] U.S.C. § 107. The U.S. fair use doctrine dates back almost 200 years. Folsom v. Marsh, 9 F. Cas. 342 (C.C.D. Mass. 1841), available at https://law.justia.com/cases/federal/district-courts/massachusetts/madce/fcas342/4104271/220/no-4.html (Last visited on June 1, 2025). However, it was not codified until 1976. See generally Matthew Sag, The Pre-History of Fair Use, 76 BROOK. L. REV. 1371 (2011) (tracing the origins of American fair use doctrine back to nineteenth-century English copyright cases on fair abridgment).

[52] See Niva Elkin-Koren, The New Frontiers of User Rights, 32 AM. U. INT'L L. REV. 1, 18–19 (2016) (tracing the Israeli fair use doctrine to the 1993 Israeli Supreme Court decision of Geva v. Walt Disney Co.).

[53] Samuelson et al., USCO Comment, supra note 11, at 15; see also Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 98 (2d Cir. 2014)

[54] Authors Guild, Inc. v. HathiTrust, 755 F. 3d 87, 97 (2d Cir. 2014) (citing Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994))

[55] [ISR.] Ministry of Just., Opinion: Uses of Copyrighted Materials For Machine Learning (2022), available at https://www.gov.il/BlobFolder/legalinfo/machine-learning/he/18-12-2022.pdf (Last visited on June 1, 2025).

[56] Ibid.

[57] Michael Geist, Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use, in The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law 157, 176 (Michael Geist ed., 2013); Ariel Katz, Fair Use 2.0: The Rebirth of Fair Dealing in Canada, in The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law (Michael Geist ed., 2013); David Vaver, User Rights: Fair Use and Beyond, 69 J. Copyright Soc'y U.S.A. 337 (2022).

[58] Ian Hargreaves, Digital Opportunity: A Review of Intellectual Property and Growth 48 (2011).

b. Express Exceptions for TDM or Computational Data Analysis

Many jurisdictions express exceptions to address the rise of TDM and computational data analysis, even without an open-ended fair use model.[59] These exceptions are especially critical in regions with a closed copyright system, where only explicitly listed uses are permitted. While more tailored than fair use, these exceptions may still require judicial assessment, such as fairness evaluations or application of international standards.[60] This approach is particularly important for AI, where the purpose is not to "enjoy" creative expression but to extract non-expressive data.

i. Japan

Japan has emerged as a frontrunner in creating a copyright exception tailored for TDM. In 2009, it became the first country to explicitly allow TDM by amending its Copyright Act.[61] A decade later, Japan expanded this exception dramatically through Article 30-4, permitting copyrighted works to be used "in any way and to the extent considered necessary" for non-expressive purposes i.e., where the user's aim is not to enjoy the work's expressive content.[62] The provision outlines acceptable activities such as data analysis, computer processing, and technological testing.

This exception, which closely resembles Germany's Freier Werkgenuss doctrine, applies to both commercial and non-commercial uses, including the training of generative AI models. However, the use must not "unreasonably prejudice the interests of the copyright owner,"[63] a clause that incorporates the latter two parts of the Berne Convention's three-step test. Alongside Article 30-4, Article 47-5 allows minor incidental uses in computerized processing, making Japan's framework the most expansive and AI-friendly globally.[64]

Nonetheless, amid growing concerns about AI, Japan has begun reassessing this openness. A May 2024 report from the Council for Cultural Affairs recommended that liability may arise for companies that knowingly use infringing datasets or circumvent technological protection measures (TPMs).[65] The report further refined what constitutes "non enjoyment" of expressive content, acknowledging the possibility of simultaneous enjoyment and non-enjoyment a subtle but essential nuance for regulating AI training.[66]

ii. United Kingdom

The United Kingdom followed the 2011 Hargreaves Review in introducing a TDM exception to promote scientific discovery.[67] In 2014, Section 29A of the Copyright, Designs and Patents Act 1988 was enacted, stating that it is not an infringement to copy a lawfully accessed work

[59] Yu, Paradigm Evolution, supra note 77, at 125 (see also Peter K. Yu, The Quest for a User-Friendly Copyright Regime in Hong Kong, 32 AM. U. INT'L L. REV. 283, 327 (2016)

[60] Ibid.

[61] Tatsuhiro Ueno, The Flexible Copyright Exception for "Non-Enjoyment" Purposes, 70 GRUR INT'L 145 (2021)

[62] Japanese Copyright Act, 1970, art. 30-4.

[63] George C. Christie, Some Key Jurisprudential Issues of the Twenty-First Century, 8 Tulane Journal of International & Comparative Law 217-232 (2000) ; Peter K. Yu, The Harmonization Game: What Basketball Can Teach About Intellectual Property and International Trade, 26 FORDHAM INT'L L.J. 218, 233–34 (2003).

[64] Japanese Copyright Act, 1970, art. 47-5.

[65] Council For Cultural Affs., Copyright Div., Subcomm. On Legal Sys., General Understanding On Ai And Copyright In Japan (2024), available in Japanese at https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoseido/r05_07/pdf/94024201_01.pdf (Last visited on June 1, 2025);

[66] ID

[67] Hargreaves Review, An Independent Review of Intellectual Property and Growth, (2011), available at https://dera.ioe.ac.uk/id/eprint/16295/7/ipreview-finalreport_Redacted.pdf (Last visited on June 1, 2025).

for non-commercial research via computational analysis.[68] This exception, however, is narrowly scoped—it only covers research with a "sole purpose" that is non-commercial, requiring lawful access, sufficient acknowledgment, and no unauthorized distribution.

Despite its narrowness, a critical advantage of the UK exception is its immunity from contractual override.[69] Even if a private contract prohibits TDM, the statutory exception takes precedence. While the UK government proposed expanding the exception to include commercial uses in 2022, the plan was abandoned in 2023, leaving researchers with limited flexibility for broader AI applications. The UK exception reflects a conservative but protective stance, safeguarding academic and nonprofit interests without fully embracing AI innovation for commercial use.[70]

iii. European Union

The Digital Single Market (DSM) Directive of 2019 established a two-tiered regime for TDM across the European Union, via Articles 3 and 4.[71] Article 3 provides a mandatory exception for research organizations and cultural heritage institutions, enabling them to carry out TDM on lawfully accessed works for scientific research, regardless of contractual terms or opt-outs.[72] This provision also allows long-term retention of data for research verification, provided security safeguards are in place.

Article 4, by contrast, provides a commercial TDM exception, open to all entities but subject to opt-out rights by rightsholders. If a copyright holder reserves rights using machine-readable means, such as metadata embedded in online content, they can prevent TDM under this provision.[73] Additionally, Article 4 allows reproductions only for as long as necessary to conduct the analysis, unlike the broader data retention permitted under Article 3.

Both articles are bounded by Article 5(5) of the InfoSoc Directive, which incorporates the three-step test: that exceptions must apply in special cases, not conflict with normal exploitation, and not unreasonably harm rightsholders.[74] A recent German case involving the LAION dataset confirmed that these exceptions apply to AI training, affirming the DSM Directive's relevance in the age of machine learning.

iv. Singapore

Singapore's 2021 Copyright Act introduced Section 244, a robust exception for computational data analysis (CDA), defined to include both traditional TDM and AI training.[75] This includes the use of works to train algorithms to identify data types, such as using images to improve

[68] Copyright, Designs and Patents Act 1988, c. 48, § 29A(1)(a).

[69] Copyright, Designs and Patents Act 1988, § 29A(5).

[70] Alina Trapova & João Pedro Quintais, The UK Government Moves Forward with a Text and Data Mining Exception for All Purposes, KLUWER COPYRIGHT BLOG (Aug. 24, 2022), https://copyrightblog.kluweriplaw.com/2022/08/24/the-ukgovernment-moves-forward-with-a-text-and-data-mining-exception-for-all-purposes. 151 See UK Withdraws Plans for Broader Text and Data Mining (TDM) Copyright and Database Right Exception, HERBERT SMITH FREEHILLS LLP (Mar. 1, 2023), https://www.herbertsmithfreehills.com/notes/ip/2023-03/uk-withdraws-plans-forbroader-text-and-data-mining-tdm-copyright-and-database-right-exception.

[71] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market, arts. 3–4, 2019 O.J. (L 130) 92

[72] Ibid.

[73] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market, arts. 2(1), 2(3), 2019 O.J. (L 130) 92.

[74] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society, art. 5(5), 2001 O.J. (L 167) 10.

[75] Copyright Act 2021 [Singapore Copyright Act] (2020 Rev Ed) § 244(1)–(2)(d), (3) (Sing.).

image recognition. Singapore's CDA exception is significant in that it explicitly applies to both commercial and non-commercial purposes, offering broad scope for innovation.[76]

However, the exception imposes key conditions: the user must have lawful access and avoid infringing sources, particularly those from "flagrantly infringing online locations."[77] The Act prohibits the circumvention of TPMs, and this exception, like those in the UK and EU Article 3, is immune from contractual override and does not require rightsholder opt-ins or consent.[78]

Uniquely, Singapore pairs this express exception with a general fair use provision, giving it a dual-track system which is a rare configuration globally.[79] This setup allows Singapore to backstop express limitations with a fair use safety net, reinforcing its strategic aim to become a regional leader in AI development. Should fair use ever be interpreted narrowly, the TDM exception ensures that innovative data uses, including AI training, remain legally protected.[80]

c. Lack of Dedicated Exceptions Despite Active AI Development

While several jurisdictions have proactively adapted their copyright laws to accommodate TDM, some leading AI nations have yet to do so. This section focuses on China and the United Arab Emirates (UAE); the two countries that have made significant strides in AI innovation but still lack dedicated legal provisions for AI training. Their experience highlights a broader trend: some countries are choosing to advance in AI development regardless of whether their copyright frameworks are fully aligned with the legal realities of large-scale model training. This contrast between technological ambition and regulatory readiness raises key questions about how law and policy shape emerging technologies.

i. China

Over the past decade, China has pursued a national strategy of AI leadership, most notably through the State Council's Next Generation Artificial Intelligence Development Plan, which aims to make China the global hub for AI innovation by 2030.[81] Chinese entities dominate in global AI indicators: according to Stanford's AI Index Report (2024), China leads the world in AI patents and industrial robot installations, with domestic tech giants like Tencent, Ping An, and Baidu, and institutions like the Chinese Academy of Sciences, filing the most patents globally for generative AI from 2014 to 2023.[82]

Yet, despite its technological momentum, China's copyright framework has not kept pace. The Third Amendment to the Copyright Law (effective June 1, 2021) maintained a closed list of limitations and exceptions, albeit with a notable update: Article 24(13) introduces a catch-all clause for "other circumstances provided for by laws and administrative regulations."[83] This

[76] Ministry of Law & Intellectual Property Office of Singapore, Singapore Copyright Review Report 32–34 (2019); Copyright Act 2021 (Singapore), No. 22 of 2021, §§ 243, 244(2)(c)(ii), 369–407.

[77] Copyright Act 2021 (Singapore), No. 22 of 2021, §§ 244(2)(d), 244(2)(e)(ii)(B), 99(1)–(2).

[78] Copyright Act 2021 (Singapore), No. 22 of 2021, §§ 187(c), 188;

[79] Section II.A.2.

[80] Peter K. Yu, The Future Path of Artificial Intelligence and Copyright Law in the Asian Pacific, 33 Mich. St. Int'l L. Rev. (forthcoming 2025)

[81] Guowuyuan Guanyu Yinfa Xinyidai Rengong Zhineng Fazhan Guihua De Tongzhi, Guofa [2017] Sanshiwu Hao Notice of the Next-Generation Artificial Intelligence Development Plan, Notice No. 35 [2017]] (issued by the State Council, July 20, 2017); Peter K. Yu, China, "Belt and Road" and Intellectual Property Cooperation, 14 GLOB. TRADE & CUSTOMS J. 244 (2019); Zhang Hongzhou & Shaleen Khanal, To Win the Great AI Race, China Turns to Southeast Asia, ASIA POL'Y, Jan. 2024, at 21.

[82] Stanford human-centered artificial intelligence, 2024 AI Index Report (2024), available at https://hai.stanford.edu/ai-index/2024-ai-index-report (Last visited on June 1, 2025).

[83] Shuimei Liu, Comments on the Chinese Fair Use Legislation, Before, In, and After 2020 (Indiana Journal of Global Legal Studies, Global Scholars Series, 2022), available at https://ijgls.indiana.edu/wp-content/uploads/2022/02/2022.02.15_

provision does not establish a fair use doctrine but opens the door for future AI-specific exceptions via implementing regulations or separate administrative measures. At present, China has not enacted a dedicated exception for AI training, but the legislative structure is now more flexible than in the past.[84]

Further complicating the picture are the Interim Measures for the Management of Generative Artificial Intelligence Services, adopted in July 2023.[85] These rules reflect China's desire to become a regulatory first-mover but their treatment of intellectual property (IP) is vague. Articles 4 and 7 mention the need to respect IP rights and avoid infringement, but they do not clarify whether AI training on copyrighted materials is legal.[86] The ultimate interpretation of these clauses will depend on judicial decisions and future regulatory action. China's challenge will be to balance state oversight with innovation freedom, particularly as geopolitical tensions with other AI powers intensify.

ii. United Arab Emirates

The UAE has made bold moves toward becoming a leader in AI, launching a national AI strategy in 2017 that aims to position the country among the top global AI nations by 2031[87]. Its open-source Falcon LLMs, developed by the Technology Innovation Institute in Abu Dhabi, have received international attention for their performance parity with models from OpenAI and Meta. Despite these developments, the UAE's Copyright and Neighboring Rights Law contains neither a fair use doctrine nor an explicit TDM exception tailored for AI training. Instead, the law offers limited scope under Article 22(1), which permits the reproduction of a single copy of a work for purely personal use and for non-profit, non-professional purposes with exclusions for computer programs, databases, and works of applied art. Although this provision may theoretically apply to AI research conducted by universities or government-linked institutions, its application to AI model training at scale remains legally unclear.[88] Moreover, the UAE's exception, like Japan's Article 30-4, is subjected to the last two parts of the Berne three-step test, meaning that any such use must not conflict with the normal exploitation of the work or unreasonably prejudice the legitimate interests of the rightsholder.

THE COST OF LEARNING: LEGAL AND TECHNICAL FAULT LINES IN UNLICENSED AI TRAINING

While the technology and jurisdictions around the world are increasingly moving in the direction of allowing AI models to be trained on proprietary content, such a trajectory may lead to a range of negative consequences that demand critical examination.[89] First, it introduces significant *technical and legal challenges*, as the process of training often involves copying and storing large volumes of copyrighted data, raising unresolved questions about infringement, unauthorized reproduction, and model opacity. Second, it has *economic and*

Shuimei-Liu_IJGLS.pdf (Last visited on June 1, 2025); see also Zhonghua Renmin Gongheguo Zhuzuoquan Fa [Copyright Law of the People's Republic of China] [2020 Chinese Copyright Law].

[84] See Yu, Long and Winding Road, (noting "the drafting of the pending implementing regulations").

[85] Shengcheng Shi Rengong Zhineng Fuwu Guanli Zhanxing Banfa [Interim Measures for the Management of Generative Artificial Intelligence Services] (promulgated by the Cyberspace Admin. of China, July 10, 2024, effective Aug. 15, 2024).

[86] Ibid.

[87] UAE Nat'l Prog. for A.I., UAE National Strategy for Artificial Intelligence 2031 7 (2018).

[88] Adam Satariano & Paul Mozur, The Global Race to Control A.I., N.Y. TIMES (Aug. 14, 2024), https//www.nytimes.com/2024/08/14/briefing/ai-china-us-technology.html (Last visited on June 1, 2025).

[89] Matthew Sag and Peter K. Yu, "The Globalization of Copyright Exceptions for AI Training," Emory Law Journal, Vol. 74, 2025 (forthcoming).

creative consequences, undermining the livelihoods of artists, writers, and other creators whose works are used without compensation, and enabling AI-generated content to serve as a market substitute for original work. Third, despite growing awareness, the *lack of transparency in training datasets* leaves rights-holders unable to verify or contest the use of their intellectual property, violating both copyright terms and database rights. Fourth, this practice has prompted *global legal and policy responses*, with countries emphasizing copyright-respecting AI systems, while others adopt more permissive stances that risk international friction. Lastly, the situation has triggered strong *industry and institutional reactions*, with publishers, artist collectives, and corporations pushing for ethical safeguards, contractual limitations, and self-regulation to prevent the unlicensed use of protected materials. These interconnected concerns form the basis of the discussion that follows.

TECHNICAL AND LEGAL CHALLENGES OF UNLICENSED TRAINING

AI training is not merely a passive process; it involves the systematic collection and analysis of large-scale datasets, which can include copyrighted material.[90] While some compare this process to reading, it is argued that it more closely resembles copying and pattern recognition.[91] Techniques such as TDM enable AI systems to detect and generalize patterns across vast corpora[92]. As mentioned in Section II, many large-scale models operate as "black boxes, and this opacity leads to high legal risk.

Researchers from Stanford and other institutions have demonstrated how models like GPT-4 can reproduce copyrighted material almost word-for-word when given the right prompt.[93] Moreover in recent times, Gen AI tools have come under increased scrutiny for replicating proprietary content without authorization. High-profile cases such as *The New York Times v. OpenAI*[94] and *Thomson Reuters v. Ross Intelligence*[95] highlights growing legal concerns over whether the use of copyrighted material in AI training and output generation constitutes infringement. In another case, Getty Images sued Stability AI [96] for allegedly training Stable Diffusion (their AI) on millions of licensed photographs without consent, pointing to the inclusion of watermarked images in outputs as evidence of direct copying.[97]

Some legal experts argue that AI developers should be treated similarly to infringers who make unauthorized reproductions.[98] As these models store compressed representations of the works they ingest, they don't merely "learn from" protected material—they internalize and potentially

[90] Copyright office, Copyright and Artificial Intelligence: Part III – The Effect of Artificial Intelligence on the Copyright Office's Practices, 7 (2024), available at https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf (Last visited on June 1, 2025).

[91] IBM, What is Big Data Analytics? available at https://www.ibm.com/think/topics/big-data-analytics (Last visited on June 1, 2025).

[92] Shubham Mukherjee, Navigating the Legalities of AI-Created Geospatial Data: An Indian Perspective, Vol. 17(2), inT. J. DigiTal earTh, 149 (2024), available at https://www.tandfonline.com/doi/full/10.1080/17538947.2024.2353122 (Last visited on June 1, 2025).

[93] Argyri Panezi, Fair Use and Artificial Intelligence: Does the Reasonable Use of Copyrighted Works Extend to AI?, Vol. 24(1), minn. J. l. sci. & Tech., 113 (2023), available at https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1564&context=mjlst (Last visited on June 1, 2025).

[94] The New York Times Co. v. Microsoft Corp. & OpenAI, Inc., No. 1:23-cv-11195 (S.D.N.Y.) (Pending).

[95] Thomson Reuters Enter. Ctr. GmbH v. Ross Intelligence Inc., No. 1:21-cv-00677 (D. Del.) (Pending).

[96] Getty Images (US), Inc. v. Stability AI, Inc., 1:23-cv-00135 (D. Del.) (Pending).

[97] Pascal G., GPT-4: The Problem Arises If These Are Your Training Materials…, P4sc4l.subsTaCK (2024), available at https://p4sc4l.substack.com/p/gpt-4-the-problem-arises-if-these (Last visited on June 1, 2025).

[98] Eleonora Rosati, The Use of Copyright Content by Generative AI Systems: What Can We Learn from the Google Books Saga? Vol. 20(3), JIPLP, 182 (2023) available at https://academic.oup.com/jiplp/article/20/3/182/7922541 (Last visited on June 1, 2025).

regurgitate it. Even when licenses are granted, the risk of breach is high.[99] If an AI system uses licensed data for broader or different purposes than allowed, it could easily exceed its authorization especially in commercial contexts where generated outputs are difficult to trace back to their origin.

ECONOMIC AND CREATIVE CONSEQUENCES

Allowing AI training on proprietary content without licensing undermines not only legal obligations but also the creative economy.[100] Creators lose direct revenue from licensing and indirect value as their work becomes training fodder for generative competitors. For instance, Canada's Standing Committee on Industry and Technology emphasized in 2023[101] that AI policy must prioritize fair remuneration for artists and warned of "economic displacement" if creators' works are freely mined.[102]A global study projects that by 2028, music and audiovisual creators could lose 24% and 21% of their revenues, respectively, due to the growing impact of generative AI on creative industries.[103]

AI companies reap substantial profits from models trained on others' creative works without paying creators. Through monetized services and enhanced features, AI companies are reaping considerable profits from models trained on copyrighted material, sparking criticism that the original authors of these works are being inadequately compensated for their contributions.[104] Scholars note that generative AI may *"unfairly compete with authors, journalists, and other creative workers"*, displacing them in the market[105] and *"threatening the livelihoods of authors, artists, and other creatives"*.[106] Publishers and news outlets across the world have alleged that AI models trained on their articles siphon away readers and ad revenue, effectively creating a "market substitute" for paywalled content.[107] Björn Ulvaeus noted that while

[99] Pranav Misra, Copyright Law in the Age of Generative AI: Reimagining the Human Authorship Principle, SSRN (2024) available at https://papers.ssrn.com/sol3/Delivery.cfm/4975857.pdf?abstractid=4975857&mirid=1 (Last visited on June 1, 2025).

[100] CISAC, Global economic study shows human creators' future at risk from generative AI, CISAC Newsroom, December 4, 2024, available at https://www.cisac.org/Newsroom/news-releases/global-economic-study-shows-human-creators-future-risk-generative-ai (Last visited on June 1, 2025).

[101] Innovation, Science and Economic Development Canada, Appearance before the Standing Committee on Industry and Technology (INDU) by the Minister of Innovation, Science and Industry - Transparency, Government of Canada, February 7, 2025, available at https://ised-isde.canada.ca/site/transparency/en/appearance-standing-committee-industry-and-technology-indu-minister-innovation-science-and-industry-0 (Last visited on June 1, 2025). House of Commons, Standing Committee on Industry and Technology (INDU), Evidence, Meeting No. 151, Parliament of Canada, March 28, 2023, available at https://www.ourcommons.ca/Content/Committee/441/INDU/Evidence/ EV13504262/INDUEV151-E.PDF (Last visited on June 1, 2025).

[102] House of Commons, Standing Committee on Industry and Technology, Evidence of the Standing Committee on Industry and Technology, Meeting No. 65, Government of Canada, March 29, 2023, available at https://publications.gc.ca/collections/collection_2023/parl/xc80-1/XC80-1-2-441-65-eng.pdf (Last visited on June 1, 2025).

[103] CISAC, Global economic study shows human creators' future at risk from generative AI, CISAC Newsroom, December 4, 2024, available at https://www.cisac.org/Newsroom/news-releases/global-economic-study-shows-human-creators-future-risk-generative-ai (Last visited on June 1, 2025).

[104] Lucchi N., ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems, Vol. 15(3), European Journal of Risk Regulation, 602-624 (2024), available at https://doi.org/10.1017/err.2023.59 (Last visited on June 1, 2025).

[105] Frank Pasquale & Haochen Sun, Consent and Compensation: Resolving Generative AI's Copyright Crisis, Virginia Law Review, April 25, 2025, available at https://virginialawreview.org/articles/consent-and-compensation-resolving-generative-ais-copyright-crisis/ (Last visited on June 1, 2025).

[106] Frank Pasquale & Haochen Sun, Consent and Compensation: Resolving Generative AI's Copyright Crisis, Virginia Law Review, April 25, 2025, available at https://virginialawreview.org/articles/consent-and-compensation-resolving-generative-ais-copyright-crisis/ (Last visited on June 1, 2025).

[107] Audrey Pope, NYT v. OpenAI: The Times's About-Face, *Harvard Law Review, April 10, 2024, available at https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/ (Last visited on June 1, 2025).; RAND*

generative AI can offer exciting opportunities for creators, poor regulation risks harming their careers and livelihoods.

AI companies rarely disclose the exact datasets used during training, making it nearly impossible for creators to determine whether their work has been ingested[108]. As a result, rights holders are left in the dark and unable to confirm if their content was used or if any opt-out requests were honored. This opacity is especially troubling given that much online content is governed by licenses or terms of use that explicitly prohibit data mining or commercial reuse.[109] Web users may only grant narrow permissions for specific purposes such as search, personal research, or educational use. Moreover, scraping large databases can also infringe sui generis database rights, which prohibit unauthorized extraction of substantial portions of a database.[110] Yet in practice, AI training often proceeds in disregard of these legal boundaries.

For instance, Getty Images has alleged that Stability AI used over 12 millions of its copyrighted images without permission to train Stable Diffusion, with the intention of directly competing with and profiting from Getty's content.[111] Similarly, the scraping of open-source code has sparked legal backlash: a class-action lawsuit against GitHub Copilot claims the tool exploits the labor of open-source developers by violating the terms of their licenses.[112] Meanwhile, AI chatbots have been caught outputting copyrighted content verbatim, including lyrics and literary passages.[113] In one case, Germany's music rights organization GEMA accused ChatGPT of reproducing protected song lyrics in response to basic prompts. These so-called hallucinations not only raise legal red flags but also risk harming the reputation and revenue streams of original creators.[114]

AI models trained on thousands of writers' texts can mimic their tone and style, offering cheap, derivative content in the marketplace. In *Authors Guild v. OpenAI,*[115] U.S. District Judge Araceli Martínez-Olguín allowed the direct copyright infringement claim to proceed, acknowledging that, if proven, OpenAI's use of copyrighted materials for training its models could constitute an unfair business practice. The plaintiffs argued that their books were used to

Corporation, Artificial Intelligence Impacts on Copyright Law, RAND Corporation, November 20, 2024, available at https://www.rand.org/pubs/perspectives/PEA3243-1.html#fn40 (Last visited on June 1, 2025).; Saikrishna & Associates, Indian Copyright Law and Generative AI: Part 3- The Output Stage: Analyzing Reproduction and Adaptation, Saikrishna & Associates, July 2024, available at https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-3-the-output-stage-analyzing-reproduction-and-adaptation/ (Last visited on June 1, 2025).

[108] J. Hardinge, E. Simperl & N. Shadbolt, We Must Fix the Lack of Transparency Around the Data Used to Train Foundation Models, Harvard Data Science Review, May 31, 2024, available at https://hdsr.mitpress.mit.edu/pub/xau9dza3 (Last visited on June 1, 2025).

[109] U.S. Copyright Office, Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication Version, U.S. Copyright Office, May 2025, available at https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf (Last visited on June 1, 2025).

[110] OECD (Organisation for Economic Co-operation and Development), Intellectual property issues in artificial intelligence trained on scraped data, OECD, February 2025, available at https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f0b/d5241a23-en.pdf (Last visited on June 1, 2025).

[111] Blake Brittain, Getty Images Lawsuit Says Stability AI Misused Photos to Train AI, Reuters, February 6, 2023, available at https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/ (Last visited on June 1, 2025).

[112] Joseph Saveri Law Firm, LLP, GitHub Copilot Intellectual Property Litigation, Joseph Saveri Law Firm, LLP, November 3, 2022, available at https://www.saverilawfirm.com/our-cases/github-copilot-intellectual-property-litigation (Last visited on June 1, 2025).

[113] OpenAI Developer Community, September 21, 2023, available at https://community.openai.com/t/some-questions-on-copyrighted-material/387428 (Last visited on June 1, 2025).

[114] GEMA, AI & Music: A Legal First - German Author and Publisher Sue Stability AI, GEMA, November 23, 2023, available at https://www.gema.de/en/news/ai-and-music/ai-lawsuit (Last visited on June 1, 2025).

[115] *Authors Guild et al. v. OpenAI Inc. et al.*, No. 1:23-cv-08292 (S.D.N.Y. filed Sept. 19, 2023).

train GPT models, allowing them to generate unauthorized sequels and summaries.[116] They argued this wasn't innovation, but "systematic theft." [117]A similar line of reasoning appeared in *Andersen v. Stability AI,*[118] where the court allowed claims to proceed against the AI firm for inducing copyright infringement by training on artists' works.[119]

Globally, the same concerns are prompting policy pushback. In Germany, the *Kneschke v. LAION*[120] case raised key questions about whether open datasets used for AI training violated photographers' rights. The Hamburg court has taken early steps, demanding greater transparency from data aggregators. Meanwhile, France's CNIL launched an investigation into OpenAI for potentially violating data protection and copyright laws—reflecting growing EU concern about the opacity and lawfulness of training methods.[121]

GLOBAL RESPONSES AND THE SHIFT TOWARD COPYRIGHT-CONSCIOUS AI

While some countries have embraced broad exceptions for information analysis others have adopted a copyright-protective stance.[122] Under Japan's 2019 amendment, even commercial mining of copyrighted works is allowed without authorization. But this laissez-faire policy has drawn sharp criticism internationally.[123] In Japan, industry groups and lawmakers have called for stronger protections and compensation mechanisms for creators, including watermarking AI-generated content and revenue-sharing models to support artists impacted by AI training. [124]

European lawmakers have taken the opposite approach. As mentioned in Section III, the EU AI Act, finalized in 2024, requires providers of general-purpose AI systems to disclose training data sources and comply with the DSM Directive.[125] The DSM grants rights-holders a reservation right even when exceptions for TDM exist, allowing them to explicitly opt out of AI training use.[126] Germany has implemented these rules rigorously, and even non-EU countries are watching closely.[127] In South Africa, for example, the Draft Intellectual Property Policy Phase II emphasizes that AI systems must not undermine creators' rights and stresses

[116] Reuters, OpenAI Gets Partial Win in Authors' U.S. Copyright Lawsuit, Reuters, February 13, 2024, available at https://www.reuters.com/legal/litigation/openai-gets-partial-win-authors-us-copyright-lawsuit-2024-02-13/ (Last visited on June 1, 2025).

[117] Hessie Jones, Generative AI is a Crisis for Copyright Law, Hessie Jones Substack, April 4, 2025, available at https://doesnotcomputeai.substack.com/p/generative-ai-and-copyright-law-a (Last visited on June 1, 2025).

[118] *Andersen et al. v. Stability AI Ltd. et al.*, No. 3:23-cv-00201 (N.D. Cal. 2024).

[119] Zach Schor, Andersen v. Stability AI: The Landmark Case Unpacking the Copyright Risks of AI Image Generators, NYU Journal of Intellectual Property & Entertainment Law, December 2, 2024, available at https://jipel.law.nyu.edu/andersen-v-stability-ai-the-landmark-case-unpacking-the-copyright-risks-of-ai-image-generators/ (Last visited on June 1, 2025).

[120] Robert Kneschke v. LAION e.V., 310 O 227/23 (Hamburg Regional Court, Germany, September 27, 2024), available at https://www.wipo.int/wipolex/en/judgments/details/2381 (Last visited on June 1, 2025).

[121] CNIL, AI and GDPR: The CNIL publishes new recommendations to support responsible innovation, CNIL, February 7, 2025, available at https://www.cnil.fr/en/ai-and-gdpr-cnil-publishes-new-recommendations-support-responsible-innovation (Last visited on June 1, 2025).

[122] Matthew Sag & Peter K. Yu, The Globalization of Copyright Exceptions for AI Training, 74 Emory L.J. 1163 (2025), available at https://ssrn.com/abstract=4976393 (Last visited on June 1, 2025).

[123] Seth Hays, AI Boom or Copyright Doom? Lessons from Asia, CEPA, March 7, 2025, available at https://cepa.org/article/ai-boom-or-copyright-doom-lessons-from-asia/ (Last visited on June 1, 2025).

[124] NAFCA, NAFCA, November 7, 2023, available at https://nafca.jp/public-comment01/ (Last visited on June 1, 2025).

[125] Artificial Intelligence Act, High-level summary of the AI Act, Artificial Intelligence Act, February 27, 2024, available at https://artificialintelligenceact.eu/high-level-summary/ (Last visited on June 1, 2025).

[126] Vaibavi S G, Balancing Innovation & Rights: A Copyright Policy Proposal for Ai Training in India, IIPRD, available at https://www.iiprd.com/balancing-innovation-rights-a-copyright-policy-proposal-for-ai-training-in-india/ (Last visited on June 1, 2025).

[127] Make it in Germany, The new Skilled Immigration Act, Make it in Germany, August 20, 2023, available at https://www.make-it-in-germany.com/en/visa-residence/skilled-immigration-act (Last visited on June 1, 2025).

the need for alignment with international copyright norms.[128] Likewise, Mexico, under the USMCA agreement, has affirmed the reproduction right and shown hesitancy to allow sweeping TDM without safeguards.[129] The country's national copyright institute has signaled that AI regulations must prevent unauthorized commercial use of protected works.

Such developments suggest a global move at least outside of a few outlier nations—toward copyright-respecting AI policy. The idea is not to halt AI progress, but to guide it within lawful and ethical frameworks. As many policymakers now emphasize, innovation must not come at the cost of expropriating the intellectual labor of others.

INDUSTRY AND INSTITUTIONAL REACTIONS

Authors' associations, media groups, and even technology firms have warned against misuse. Penguin Random House has included new clauses in contracts to prohibit AI training on their books. Thousands of creatives across the UK, Australia, and the EU have signed letters demanding that generative AI systems respect licensing. In India, the Indian Performing Right Society (IPRS) has expressed concern that AI-generated music trained on copyrighted tracks might violate the rights of composers and performers.

Meanwhile, companies like Wipro have embedded AI guidelines that restrict training on sensitive or proprietary data, noting the copyright and reputational risks. In a global business environment increasingly shaped by ethical AI principles, such corporate self-regulation aligns with the legal direction many governments are already taking.

INDIA AT THE CROSSROAD: A POLICY BLUEPRINT FOR INDIA

After analyzing the position of Copyright with regard to its training process throughout the world, it is pertinent to look at the Indian position in the same regard. As one of the largest creators and consumers of digital content globally, and a rising power in AI research and deployment, the country must shape a policy landscape that reflects both its technological ambitions and its foundational respect for creators' rights.[130] The way forward lies in striking a careful balance embracing innovation while ensuring that the intellectual and creative labor fueling AI systems is not stripped of value or recognition.[131]

One of the core challenges in India is that its current legal and regulatory infrastructure is not fully equipped to handle the complexities posed by generative AI.[132] The Copyright Act, while comprehensive for traditional media, does not clearly define how large-scale text and data mining or deep learning processes interact with exclusive rights like reproduction or adaptation. In 2020, the Indian Copyright Office controversially granted registration to a work of art titled

---

[128] Dane Bottomley, A Comparative Analysis of the need for sui generis Artificial Intelligence Legislation in Kenya and South Africa, Centre for Intellectual Property and Information Technology Law (CIPIT), Strathmore University, February 2024, available at https://cipit.strathmore.edu/wp-content/uploads/2024/09/A-Comparative-Analysis-of-the-need-for-sui-generis-Artificial-Intelligence.pdf (Last visited on June 1, 2025).

[129] Office of the United States Trade Representative, United States-Mexico-Canada Agreement, United States Trade Representative (USTR), July 1, 2020, available at https://ustr.gov/trade-agreements/free-trade-agreements/united-states-mexico-canada-agreement (Last visited on June 1, 2025).

[130] NITI Aayog, National Strategy for Artificial Intelligence #AIForAll, Government of India, June 2018, available at https://www.niti.gov.in/sites/default/files/2021-07/NationalStrategy-for-AI-Discussion-Paper.pdf (Last visited on June 1, 2025).

[131] WIPO, WIPO Technology Trends 2019: Artificial Intelligence, World Intellectual Property Organization, 2019, available at https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf (Last visited on June 1, 2025).

[132] Arpan Gupta, AI and the Issue of Human-centricity in Copyright Law, SSRN, 2024, available at https://ssrn.com/abstract=4987158 (Last visited on June 1, 2025).

Suryast, naming an AI system, RAGHAV, as a co-author and despite no legislative amendment to the Copyright Act, 1957 permitting non-human authorship.[133] Although the office later issued a withdrawal notice, the registration still appears valid on public records, reflecting the legal ambiguity and institutional hesitation surrounding AI authorship in India.[134]

Unlike the U.S., India has neither issued clear disclosure norms for AI-generated works nor launched public consultations to address this growing challenge. While the 161st Parliamentary Standing Committee Report urged reforms in copyright and patent law to incorporate AI-related innovations, its recommendations lacked empirical grounding in the practical needs and risks of India's AI ecosystem.[135] In the absence of clarification, this ambiguity risks being interpreted to the advantage of large technology players, particularly those capable of scraping vast quantities of content under vague claims of fair dealing.

What India requires is a policy framework that formally recognizes the role of consent and compensation within the AI training pipeline, while simultaneously encouraging the creation and use of open datasets that do not infringe upon proprietary rights. The utilization of copyrighted works, particularly for commercial model training, raises important legal and ethical considerations that merit the establishment of clear authorization mechanisms. One viable approach may involve the adoption of a licensing regime, ideally facilitated through collective rights management structures[136] to provide an efficient and scalable framework for lawful access. Analogous to the copyright societies in the music industry, such a model could enable equitable remuneration and attribution for authors, journalists, filmmakers, and visual artists whose creative outputs form the basis of training data for generative AI systems.

At the same time, Government-funded AI initiatives can focus on building corpora composed of public domain materials,[137] official documents, and voluntarily submitted creative content. India's multilingual richness offers a unique advantage here: curating training datasets in regional languages not only boosts linguistic diversity in AI outputs but does so without entangling legal risk.[138] A key component of the policy should also include greater transparency obligations for AI developers operating in India. This could take the form of mandatory disclosures about what datasets were used for training, whether licenses were obtained, and what measures are in place to prevent the model from producing infringing outputs. These disclosures need not stifle innovation; rather, they would instill a culture of accountability that benefits both creators and end-users. Countries like Canada and members

[133] Copyright Office, Government of India, Public Register Entry: Work Titled "Suryast" [CO/A/2020/5562], Indian Copyright Office, 2020, available at https://copyright.gov.in/PublicSearch.aspx (Last visited on June 1, 2025).

[134] Areeb Uddin Ahmed, AI As Author? Indian Copyright Office's Withdrawal of Registration For AI-Generated Artwork Sparks Debate, LiveLaw, December 16, 2022, available at https://www.livelaw.in/news-updates/ai-as-author-indian-copyright-office-withdraws-registration-217329 (Last visited on June 1, 2025).

[135] Department-related Parliamentary Standing Committee on Commerce, 161st Report on the Review of the Intellectual Property Rights Regime in India, Rajya Sabha Secretariat, 2021, available at https://rajyasabha.nic.in/rsnew/committee_pages/Commerce_161.pdf (Last visited on June 1, 2025).

[136] M. Kretschmer & P. Towse (eds), Collective Management of Copyright and Related Rights (2nd ed., Springer 2020).

[137] OECD, Encouraging Responsible Data Sharing and Use of Open Datasets for AI Training, OECD, December 12, 2022, available at https://oecd.ai/en/data-sharing (Last visited on June 1, 2025).

[138] Luciano Floridi et al., AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, 28(4) Minds and Machines 689 (2018), available at https://link.springer.com/article/10.1007/s11023-018-9486-5 (La st visited on June 1, 2025).

of the European Union are already implementing such requirements and India should not fall behind.[139]

Policymakers also need to be attentive to the broader economic context. India's creative economy employs millions. If these sectors are undercut by AI tools that exploit their content without return, the damage will extend beyond lost royalties. It will erode trust, discourage investment in creative careers, and potentially lead to cultural homogenization as human creators struggle to compete with algorithmic derivatives. A responsible policy can prevent this by ensuring AI functions as a collaborator, not a usurper. India's Copyright Act of 1957, though periodically amended, still largely reflects a pre-digital, pre-AI sensibility. Its focus remains on regulating individual acts of copying or distribution, not the complex computational reproduction and synthesis that AI systems now perform. While Section 52 provides limited fair dealing exceptions, the Act lacks a coherent framework to address large-scale, automated extraction and abstraction of creative content through machine learning.[140] However, rather than overhauling the entire Act, India has a workable opportunity to modernize its copyright law by embedding AI-specific clarity through nuanced amendments. Some potential changes that can be made in the Act are:

Section 2[141] of the Act which defines terms such as "reproduction," "copy," and "communication to the public" could be expanded to clarify that algorithmic ingestion of copyrighted content for the purpose of model training constitutes a form of reproduction or adaptation. This would close a critical loophole that AI developers may otherwise exploit.

Similarly, Section 14[142] which enumerates the exclusive rights of authors, can be amended to recognize the right to license or restrict the use of their works for machine learning purposes. This mirrors provisions in the EU's DSM Directive and France's Code of Intellectual Property, where authors retain explicit reservation rights over the TDM of their content. Such amendments would allow training on protected works in exchange for royalties and registration, monitored by a specialized regulator such as the proposed Copyright and AI Regulatory Authority (CARA). CARA could serve this function by developing standard contract templates, managing collective licensing schemes, and enforcing the limits of lawful use.

The Act's licensing provisions under Sections 30–32[143] could also be adapted to include statutory or compulsory licenses specifically for AI training could empower authors to assert their attribution rights even against derivative AI outputs, supported by watermarking technologies or metadata standards embedded in training protocols.

India should also consider inserting a new chapter in the Copyright Act dedicated to digital and algorithmic usage analogous to Chapter IX of the IT Act, which deals with electronic records and cyber offenses. This new chapter could define key terms such as "machine learning," "generative output," and "training data," and outline compliance requirements for AI developers. These could include obligations to disclose data provenance, maintain audit logs,

[139] Government of Canada, Artificial Intelligence and Data Act (AIDA), Part of Bill C-27, Innovation, Science and Economic Development Canada, June 2022, available at https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act (Last visited on June 1, 2025).

[140] Copyright Act, 1957, § 52.

[141] Copyright Act, 1957, § 2.

[142] Copyright Act, 1957, § 14.

[143] Copyright Act, 1957, §§ 30-32.

and respect opt-outs from creators. This aligns with international discussions at WIPO, which is currently exploring how national IP systems can adapt to AI technologies.

Importantly, the need for amendment is not theoretical but it is driven by rapid developments in the Indian AI ecosystem. As generative models become integral to government e-governance tools, language translation engines, and edtech platforms, the risk of systemic copyright infringement increases. Without a legal framework that defines boundaries, creators particularly from non-English, regional, and indigenous traditions face the prospect of silent, large-scale appropriation.

CONCLUSION

As artificial intelligence continues to redefine the boundaries of creation, cognition, and commerce, copyright law finds itself at a historical inflection point. The core premise of traditional copyright that protection flows from human creativity is now confronted by a paradigm in which machines absorb, analyze, and reproduce expressive works at unprecedented scale and opacity. From tokenization to deep learning, the training process of generative AI introduces legal challenges that existing frameworks were never designed to address. The central concern is not just about copying but it is about the silent, large-scale abstraction of creative labor and the uncertain impact it has on authorship, attribution, and market fairness. Global responses to this dilemma have been diverse. While jurisdictions like the United States, Israel, and Singapore have embraced flexibility through fair use or express exceptions for text and data mining, others such as the European Union have taken a more cautious, copyright-conscious stance. Meanwhile, countries like China and the UAE surge ahead in AI development despite lacking tailored copyright regulations, creating a regulatory vacuum with long-term risks. For India, the path forward demands legal clarity, institutional readiness, and cultural sensitivity. A revised framework must affirm that innovation does not justify appropriation. It should prioritize consent, licensing, and transparency while promoting open access to non-proprietary training data. Crucially, India must guard against systemic devaluation of its creative economy, ensuring that AI serves as a collaborator of human ingenuity. Through pragmatic reforms, India has the opportunity to pioneer a balanced, future-ready copyright regime that protects creators, empowers innovation, and upholds the integrity of its rich intellectual heritage.