EU AI ACT, TRAINING DATA, AND MUSIC COPYRIGHT: POLICY LESSONS FOR INDIA

(ADITYA SHARMA) [1]

ABSTRACT

Artificial Intelligence (AI) has started to profoundly reshape the music industry by generating compositions, imitating performers, and transforming modes of distribution and consumption. While these developments increase creative possibilities, they also disrupt settled understandings of authorship, originality, and remuneration in copyright law. A central legal concern is the use of extensive quantities of copyrighted musical works as training data for generative AI models, often without authorization or compensation. The European Union's Artificial Intelligence Act (EU AI Act), adopted in 2024, introduces a pioneering response by imposing transparency obligations on providers of generative AI systems, which includes disclosure of the datasets used for training. This mandate has immediate implications for music copyright and it may empower rightsholders to detect unauthorized uses, facilitate licensing arrangements, and promote accountability across the value chain which will affect the corporate governance of the company as well.

At the same time, the Act raises rather difficult questions about enforceability, the scope of disclosure duties, and the risk that excessive compliance burdens may chill technological innovation. Whilst, the Indian Copyright Act, 1957, while robust in its protection of authors and performers, does not address the unauthorized use of works in AI development or the status of AI-generated music. As a result, Indian courts and policymakers don't have clear tools to address conflicts between innovation and rights protection in this emerging field.

Drawing lessons from the EU, this paper argues that India should consider adopting a calibrated framework and potentially a disclosure-based or statutory licensing mechanisms to ensure lawful AI development as well as safeguarding the interests of artists and the music industry.

Keywords: EU AI Act, music copyright, training data, generative AI, India, statutory licensing.

[1] Advocate in Supreme Court of India and Delhi High Court.

I.INTRODUCTION

Artificial Intelligence (AI) is rapidly transmuting creative industries, particularly the music sector, by enabling machines to compose songs, replicate vocal styles, and generate instrumental tracks at scale. While these technologies develop artistic possibilities and democratise access to music production, they simultaneously unsettle long-standing foundations of copyright law, including originality, authorship, and remuneration of creators. The global debate has now crystallised around one important issue: the use of copyrighted works as *training data* for generative AI models, often without consent or compensation to rightsholders.

The European Union has taken the first substantial regulatory step with the adoption of the Artificial Intelligence Act (EU AI Act) in 2024, which categorises AI systems by risk and, importantly for copyright, requires generative AI developers to disclose information about the training datasets used.[2] This transparency obligation, although procedural in appearance, has considerable implications for the music industry as it may allow rights holders to trace whether their works were used without authorisation, thereby enabling enforcement or licensing negotiations.

Beyond the realm of copyright, the corporate law dimension of generative AI demands equal attention. The development and commercialisation of AI systems for music creation are largely driven by corporate entities, from global technology companies to domestic streaming platforms, whose governance choices and compliance frameworks will determine how copyright obligations are observed in practice. Questions of board-level accountability, directors' duties of care, and corporate disclosure overlap with the ethical and lawful deployment of AI tools. Moreover, issues of competition law, mergers and acquisitions in the music-tech sector, and investor due diligence over the legality of training datasets illustrate that AI-driven disruption in music is concurrently a matter of intellectual property and corporate responsibility.[3]

India, by contrast, finds itself at a regulatory intersection. The Copyright Act, 1957, while robust in protecting literary, dramatic, musical, and artistic works, was conceived in a pre-

[2] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 12 July 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 212/1, art 53.
[3] Competition Act 2002 (India), section 4.

digital era and remains silent on both the training of AI models using copyrighted works and the legal status of AI-generated yields.[4] Courts and policymakers thus lack a clear legal framework to resolve disputes arising from AI's use of copyrighted material, raising essential questions about innovation, fairness, and the safety of cultural industries.

This paper begins by situating the issue of AI and training data within the broader trajectory of copyright law, before analysing the EU's regulatory response under the AI Act and its implications for music copyright. It then considers the Indian legal landscape, highlighting doctrinal gaps and policy challenges, and argues for a calibrated framework that balances technological innovation with the protection of authors' rights. Ultimately, India must draw lessons from the EU's practical approach while tailoring solutions to its unique socio-economic and cultural context.

II. BACKGROUND – AI, TRAINING DATA, AND COPYRIGHT LAW

The development of generative AI systems relies majorly on large-scale datasets containing vast quantities of creative works, including music, lyrics, and sound recordings. These datasets are used to "train" models to recognise stylistic patterns, harmonic progressions, and vocal timbres, enabling them to generate outputs that often seem like human-created compositions. The use of such training data has raised essential questions under copyright law, particularly regarding whether ingesting copyrighted works without licence constitutes infringement.

In copyright jurisprudence, a crucial difference lies between the *use* of works for "analysis" or "learning" and the *reproduction* of shielded expression in derivative outputs. Courts in the United States, for instance, in cases like *Authors Guild v Google*, held that digitising books for the purpose of creating a searchable index could qualify as fair use because it was transformative and provided a socially beneficial function. By analogy, AI developers claim that training on copyrighted works is a non-expressive use parallel to analysis, and therefore should not be regarded as infringement.[5]

The European Union has taken a more watchful stance. In *Infopaq International A/S v Danske Dagblades Forening*, the Court of Justice of the European Union (CJEU) held that even the

[4] Copyright Act, 1957 (India), section 13–14.
[5] *Authors Guild v Google Inc* 804 F 3d 202 (2d Cir 2015).

reproduction of eleven words from a newspaper article could amount to a reproduction "in part" under the Copyright Directive.[6] This expansive view of reproduction suggests that unauthorised text or data mining of creative works may, in principle, infringe copyright unless expressly exempted. The EU has since sought to harmonise these tensions by embracing provisions under the Directive on Copyright in the Digital Single Market (DSM Directive), which introduced limited text and data mining exceptions.[7]

India's copyright jurisprudence has not yet directly confronted the question of AI training. The Copyright Act 1957 provides strong protections for "musical works" and "sound recordings" but lacks exceptions comparable to fair use or text and data mining.[8] This leaves a regulatory vacuum where AI developers could be exposed to liability for focusing on copyrighted works without licences, while creators lack clarity on whether their rights extend to such uses.

Thus, the international experience demonstrates two contending trajectories: the US emphasis on flexibility and fair use, and the EU emphasis on strict protection balanced by narrow statutory exceptions. India, hovering between these models, must decide how to regulate the use of copyrighted works in AI development while safeguarding its cultural industries.

III. THE EU AI ACT AND ITS IMPLICATIONS FOR MUSIC COPYRIGHT

The European Union's Artificial Intelligence Act (EU AI Act), adopted in 2024, represents the first comprehensive attempt to standardize artificial intelligence across sectors.[9] Its framework is primarily risk-based, categorising AI systems as "unacceptable risk," "high risk," and "limited risk." Generative AI models, such as those capable of producing music, fall within the concluding categories and are subject to transparency obligations rather than outright bans.

For the music industry, the most significant provision is Article 53, which obligates providers of general-purpose AI models to disclose "a sufficiently detailed summary of the content used for training." This marks a crucial point in copyright enforcement because it enables rightsholders to scrutinise whether their works were used in training datasets. In theory, such

[6] Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECR I-6569.
[7] 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 [2019] OJ L 130/92, arts 3–4.
[8] Copyright Act 1957 (India), s 2(p), s 13(1)(a)–(c).
[9] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 12 July 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 212/1.

disclosure could empower composers, lyricists, and record labels to pursue remedies for unauthorised uses, demand licensing arrangements, or negotiate through collective remuneration schemes.

However, the Act stops short of obliging developers to publish full datasets or provide individualised notifications to rightsholders. The obligation is framed in broad terms, leaving significant discretion to AI providers as to the level of detail divulged. Critics argue that vague disclosure standards may dent enforceability, since rightsholders cannot realistically identify infringements without granular information. Conversely, industry advocates warn that unwarranted disclosure could reveal trade secrets, impose prohibitive compliance costs, and discourage innovation in Europe relative to jurisdictions like the United States or China.[10]

From a copyright perspective, the EU AI Act can be apprehended as creating a procedural bridge between AI innovation and rights protection. Instead of directly outlawing the use of copyrighted works as training data, the Act pioneers transparency measures that facilitate negotiation and enforcement. This aligns with the broader EU approach of promoting accountability in digital markets, exemplified by parallel legislation such as the Digital Services Act and the Digital Markets Act, which also rely on transparency and reporting obligations to regulate complex technological ecosystems.[11]

The inferences for music copyright are thus twofold. First, the Act increases the bargaining power of rightsholders by requiring disclosures that may reveal unauthorised training practices. Second, it opens the door to policy innovations such as collective licensing schemes or extended statutory licences that could resolve the mass use of musical works in AI training with fair remuneration for creators. India, seeking to modernise its copyright regime, can draw valuable lessons from this existing approach, which avoids prohibiting innovation but ensures transparency and accountability in AI development.

[10] European Parliament, 'Debate on the Artificial Intelligence Act' (Plenary Sitting, 13 March 2024) https://www.europarl.europa.eu accessed 15 September 2025.
[11] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act) [2022] OJ L 277/1.

IV. THE INDIAN COPYRIGHT FRAMEWORK AND AI-GAPS AND CHALLENGES

The Indian copyright regime is essentially governed by the Copyright Act 1957, which extends protection to literary, dramatic, musical, and artistic works, alongside sound recordings and cinematographic films.[12] The Act has undergone periodic amendments, especially in 1994 and 2012, to address digital technologies and performers' rights. Yet it remains anchored in a human-centric perception of authorship and originality.

Under Section 2(d), an "author" is defined in relation to the type of oeuvre, such as the composer in the case of a musical work or the producer in the case of a sound recording.[13] Nowhere does the statute recognise AI systems or their developers as potential authors. As a result, AI-generated musical outputs majorly fall into a legal grey zone: if human input is minimal, they may not qualify for copyright protection at all; if human input is substantial, ownership will vest in the human author, not the AI system. This uncertainty mirrors international debates but is particularly pressing in India, where courts have historically adhered to a strict interpretation of originality.

Indian jurisprudence is having this human-centred approach. In *Eastern Book Company v D B Modak*, the Supreme Court held that copyright mandates "a minimum degree of creativity" rather than mere labour or investment.[14] Applying this touchstone to AI outputs suggests that machine-generated works without human creative control are unlikely to be protected. Conversely, training AI on copyrighted works without authorisation raises questions of infringement, yet the Act provides no overt guidance on whether such uses amount to reproduction, adaptation, or fair dealing.

The Act's fair dealing provision under Section 52 is restricted compared to the US doctrine of fair use. It permits limited use of works for tenacities such as private study, criticism, or reporting current events. Unlike the EU's DSM Directive, Indian law does not give exceptions for text and data mining (TDM), which has emerged as the legal foundation for AI training in Europe. This omission leaves Indian developers in a perilous position: large-scale AI training

[12] Copyright Act 1957 (India), s 13(1).
[13] Copyright Act 1957 (India), s 2(d).
[14] *Eastern Book Company v D B Modak* (2008) 1 SCC 1 (15).

using copyrighted music could expose them to liability, while the absence of statutory licensing schemes denies rightsholders a structured means of remuneration.

Policymakers also face some practical enforcement challenges. Detecting whether copyrighted works have been used in AI training is technically challenging, especially without transparency obligations similar to those in the EU AI Act. Further, India's execution infrastructure already struggles with widespread piracy and unauthorised digital distribution.
Adding AI-related disputes without clear statutory guidance risks developing legal uncertainty for both innovators and creators.

Thus, India confronts a two-fold challenge: first, simplifying the status of AI-generated works under its originality and authorship doctrines; and second, designing a regulatory mechanism to govern the use of copyrighted works in AI training. Without reform, India risks crawling behind global trends and exposing its creative industries to unremunerated exploitation.

V. POLICY OPTIONS FOR INDIA – DISCLOSURE, LICENSING, AND BEYOND

India faces an imminent need to modernise its copyright framework to address the difficulties posed by AI training and generative music. The status quo is where AI training is neither expressly permitted nor regulated and creates ambiguity that undermines both innovation and the protection of creators. By examining international models, particularly the EU AI Act and the DSM Directive, India can develop a highly attuned response. Three policy avenues stand out of all which are: transparency obligations, statutory licensing, and collective rights management.

1. TRANSPARENCY AND DISCLOSURE OBLIGATIONS

One of the most valuable lessons from the EU AI Act is the role of discovery in enabling accountability. Article 53 requires providers of general-purpose AI models to publish "a sufficiently detailed summary" of all of their training data.[15] For India, a similar obligation could be integrated into either the Copyright Act or a discrete AI governance framework.

[15] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 12 July 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 212/1, art 53.

Even a high-level disclosure which will identifying categories of works (e.g., "Indian classical music archives" or "Bollywood recordings from 1980–2000") can encourage rightsholders to monitor potential unauthorised uses.

Such a reform would not only protect artists but also cultivate trust in AI development. However, policymakers must carefully balance discovery with trade secret protections. Excessive transparency could deter investment in AI research, a disquiet also voiced during debates in the European Parliament.[16] A tailored approach, requiring summaries rather than raw datasets, looks more feasible for India.

2. STATUTORY LICENSING SCHEMES

India is already having experience with statutory licensing in the music industry. Section 31D of the Copyright Act allows broadcasting organisations to use sound recordings by paying royalties at rates fixed by the Copyright Board in India.[17] This mechanism could serve as a template for AI training and developers could access copyrighted works under a compulsory or statutory licence, provided they pay equitable remuneration to a collective rights management body.

Such a scheme would address two pressing concerns which are the impracticality of negotiating individual licences for millions of works, and the need to ensure fair compensation for artists. A statutory licence for AI training would correspond to international proposals for "extended collective licensing" in the digital environment.[18]

3. COLLECTIVE RIGHTS MANAGEMENT AND SECTORAL GUIDELINES

Given the fragmented nature of India's music industry, effective implementation may require toughening collective management organisations (CMOs) such as the Indian Performing Right Society (IPRS). CMOs could play a central role in navigating blanket licences with AI developers, distributing royalties, and ensuring compliance. This approach has parallels in the EU, where collective management has long enabled efficient licensing in music and

[16] European Parliament, 'Debate on the Artificial Intelligence Act' (Plenary Sitting, 13 March 2024) https://www.europarl.europa.eu accessed 18 September 2025.
[17] Copyright Act 1957 (India), s 31D.
[18] Lucie Guibault, 'Collective Management of Copyright and Related Rights in the Digital Single Market' (2019) 50 IIC 547, 562.

broadcasting.[19] In addition, India could foresee sectoral guidelines for AI developers, outlining best practices on dataset curation, attribution of sources, and royalty distribution. Such soft-law instruments could serve as an interim measure while a comprehensive legislative reform is debated.

4. BALANCING INNOVATION AND RIGHTS PROTECTION

Finally, India's reforms must remain sensitive to broader policy goals, including nurturing innovation, promoting access to technology, and supporting cultural industries. Heavy-handed regulation may risk stifling emerging AI research, while unchecked exploitation of copyrighted works would undermine the livelihood of artists. The optimal approach lies in calibrated regulation which is combining limited transparency duties with statutory licensing mechanisms that distribute benefits equitably.

While the copyright dimensions of AI-generated music dominate majority of debates, the corporate law perspective is equally crucial. Issues of governance, liability, market concentration, and investment all suggest that generative AI cannot be structured in silos. Corporate boards will gradually face the responsibility of ensuring lawful use of AI technologies, while M&A transactions, investor due diligence, and ESG disclosures will bring copyright compliance to the core of corporate law practice in the digital era.

[19] Daniel Gervais, 'Collective Management of Copyright: Theory and Practice in the Digital Age' (2nd edn, Kluwer Law International 2016) 215.

REFERENCES

1. Authors Guild v. Google Inc., 804 F.3d 202 (2d Cir. 2015).
2. Companies Act, 2013 (India), s. 166(2).
3. Competition Act, 2002 (India), s. 4.
4. Copyright Act, 1957 (India), ss. 2(d), 2(p), 13, 14, 31D, 52.
5. 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 [2019] OJ L 130/92.
6. Eastern Book Company v. D.B. Modak, (2008) 1 SCC 1.
7. European Parliament, "Debate on the Artificial Intelligence Act", Plenary Sitting, 13 March 2024, available at https://www.europarl.europa.eu.
8. European Parliament, "Debate on the Artificial Intelligence Act", Plenary Sitting, 18 September 2025, available at https://www.europarl.europa.eu.
9. Gervais, Daniel J., *Collective Management of Copyright: Theory and Practice in the Digital Age* (2nd ed., Kluwer Law International, 2016).
10. Guibault, Lucie, "Collective Management of Copyright and Related Rights in the Digital Single Market", 50 *International Review of Intellectual Property and Competition Law (IIC)* 547 (2019).
11. Infopaq International A/S v. Danske Dagblades Forening, Case C-5/08, [2009] ECR I-6569.
12. Keller, Perry & Aplin, Tanya F., "Reconciling Trade Secrets and AI Public Transparency" (SSRN Working Paper, 4 December 2024), available at https://ssrn.com/abstract=5044278.
13. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act) [2022] OJ L 277/1.
14. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 12 July 2024 laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) [2024] OJ L 212/1.