Page: 1670-1687 *URL: www.lijdlr.com*

LAWFOYER INTERNATIONAL JOURNAL OF DOCTRINAL LEGAL RESEARCH

ISSN: 2583-7753

LawFoyer International Journal of Doctrinal Legal Research

[ISSN: 2583-7753]

Volume 3 | Issue 4

2025

DOI: https://doi.org/10.70183/lijdlr.2025.v03.194

CHALLENGES AND SOLUTIONS FOR COPYRIGHT INFRINGEMENT IN THE DIGITAL AGE

Dhananjaya C[1]

I. ABSTRACT

This paper examines the adequacy of India's copyright and intermediary liability framework in responding to the challenges posed by algorithmically driven digital platforms and the use of copyrighted material in artificial intelligence training datasets. The research problem arises from the growing disjunction between the reactive, notice-based regime under the Copyright Act, 1957 and the Information Technology Act, 2000, and the systemic realities of content curation, recommendation and large-scale machine learning. While judicial decisions such as Eastern Book Company v. D. B. Modak and MySpace Inc. v. Super Cassettes Industries Ltd have clarified originality standards and the contours of "actual knowledge", they do not resolve whether intermediaries that algorithmically promote infringing content retain safe-harbour protection, nor whether the ingestion of protected works for AI model training constitutes infringement under Indian law. This study adopts a doctrinal and comparative research methodology. Primary sources include Indian statutes, rules and leading judicial precedents, while secondary sources comprise academic literature, policy reports and comparative materials from the European Union, the United States and Germany. Through analytical synthesis, the paper evaluates the effectiveness of India's current notice-and-takedown framework and assesses the suitability of foreign models such as the EU's text and data mining exceptions and collective licensing mechanisms. The key finding is that Indian law remains structurally ill-equipped to regulate algorithmic promotion of infringing content and AI-training datasets. The absence of statutory clarity creates uncertainty for creators, intermediaries and developers alike. The paper concludes by recommending a calibrated reform package including clearer standards of "actual knowledge", mandatory transparency obligations for large platforms, and a statutory or collective licensing framework for AI training datasets. These measures aim to balance effective copyright enforcement with constitutional values of free expression and innovation.

[1] 9th semester student pursuing B.B.A LL.B at KLE Law College, Bengaluru, Karnataka (India). Email: dhananjayac736@gmail.com

II. KEYWORDS

Intermediary liability, Algorithmic platforms, Copyright enforcement, Artificial intelligence training, Text and data mining

III. INTRODUCTION

The digital revolution has fundamentally altered how creative works are produced, distributed and consumed. What once required physical duplication can now be copied, indexed and redistributed worldwide in seconds; platforms, caches and machine-learning datasets intermediate nearly every stage of creation and consumption. This transformation has exposed two enduring tensions for copyright law. First, many doctrines originality, authorship and territorial enforcement were conceived for an analogue era and strain under cross-border hosting, algorithmic curation and large-scale text and data mining. Secondly, the rise of technological enforcement measures such as digital rights management (DRM), content-ID systems and automated filtering has created difficult trade-offs between efficient rights protection and the preservation of legitimate expression, due process and open innovation.[2]

India's legal framework is anchored in the Copyright Act, 1957, which sets out exclusive rights, exceptions and statutory licensing. It is complemented by the Information Technology Act, 2000 and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, which provide a conditional safe harbour for intermediaries.[3] Indian case law has already begun to wrestle with how responsibility should be allocated between platforms and users. The Supreme Court's decision in *Eastern Book Company v D. B. Modak* moved Indian originality doctrine towards a "modicum of creativity" test, while the Delhi High Court's

[2] Martin Husovec et al, 'The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown?' (SSRN, 2020).
[3] Information Technology Act 2000 (India) s 79; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (India).

 (ISSN: 2583-7753)

decision in *Super Cassettes Industries Ltd v MySpace Inc* dealt with intermediary liability for user-generated content and the triggers for takedown duties.[4]

More recently, generative AI and large-scale model training have created a new front. In November 2024, Indian news agency ANI filed suit against OpenAI in the Delhi High Court alleging unauthorized use of its copyrighted materials for model training the first high-profile Indian litigation on AI training datasets.[5] This case highlights the absence of clear guidance in Indian law on whether the ingestion of protected works for machine learning constitutes infringement or falls within existing exceptions.

A. Research Objectives

1. To examine the adequacy of the existing Indian legal framework governing intermediary liability in the context of algorithmic recommendation and promotion of infringing content.
2. To analyse the copyright implications of using protected works in AI training datasets under Indian law.
3. To undertake a comparative assessment of regulatory approaches in the European Union, United States and Germany with respect to intermediary obligations and text and data mining exceptions.
4. To propose calibrated reforms suitable for the Indian constitutional and statutory framework.

B. Research Hypotheses

1. Clearer statutory standards defining "actual knowledge" in relation to algorithmic promotion will improve copyright enforcement without chilling legitimate digital innovation.

[4] Eastern Book Company & Ors v D. B. Modak & Anr AIR 2008 SC 809; Super Cassettes Industries Ltd v MySpace Inc SCC Online Del 6382 (Delhi High Court, 23 December 2016).
[5] ANI Media Pvt Ltd v OpenAI Inc & Anr CS(COMM) 1028/2024 (Delhi High Court); Reuters, 'Indian news agency ANI sues OpenAI for unsanctioned content use in AI training' (19 November 2024) https://www.reuters.com.

2. The absence of a statutory licensing or exception framework for AI training datasets significantly increases legal uncertainty for developers and rightsholders in India.
3. Introducing collective or compulsory licensing mechanisms for AI training will provide more efficient remuneration to creators than a purely litigation-driven model.

C. Research Methodology

This study adopts a doctrinal or black-letter legal research methodology supplemented by comparative analysis. Primary data sources include the Copyright Act, 1957, the Information Technology Act, 2000, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, and leading Indian judicial decisions such as *Eastern Book Company v. D. B. Modak* and *MySpace Inc. v. Super Cassettes Industries Ltd*. Secondary sources comprise scholarly journal articles, policy reports, international instruments and comparative materials from the European Union, United States and Germany.

The analytical framework involves identifying doctrinal gaps in Indian law, evaluating judicial reasoning on intermediary liability and originality, and synthesising comparative regulatory models on text and data mining and platform liability. This approach enables the formulation of normative recommendations grounded in both Indian constitutional principles and international best practices.

D. Conceptual Background

Copyright is a legal regime designed to incentivise the creation and dissemination of original works by granting authors a bundle of exclusive rights over reproduction, communication and adaptation.[6] In India, the framework is anchored in the Copyright Act, 1957, which has been periodically amended to align with technological advances and international obligations under the Berne Convention and the TRIPS Agreement.[7]

[6] Lionel Bentley and Brad Sherman, Intellectual Property Law (6th edn, OUP 2022).
[7] Copyright Act 1957 (India); Berne Convention for the Protection of Literary and Artistic Works (1971); Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), Marrakesh Agreement Establishing the World Trade Organization, Annex 1C (1994).

The Digital Age has fundamentally altered the conditions under which copyright operates. Works can be reproduced and distributed instantaneously across jurisdictions; intermediaries and platforms curate and monetise content at massive scale; and machine-learning systems ingest vast corpora of text, images and audio to generate new outputs. These developments have blurred the boundaries between user, intermediary and author, and have strained doctrines developed for an analogue environment, including originality, authorship, and secondary liability.[8]

Two specific challenges stand out in the Indian context. First, the conditional safe harbour for intermediaries under Section 79 of the Information Technology Act does not expressly address algorithmic recommendation or promotion of infringing content. Secondly, Indian law has yet to clarify the copyright implications of using protected works in AI training datasets. These challenges, which form the research questions for this study, illustrate the need for a reassessment of existing doctrines and for calibrated reforms that balance effective enforcement with the constitutional guarantee of freedom of expression and the promotion of innovation.[9]

E. Significance of Study

This study is significant for three interlinked reasons:

1. It addresses an urgent doctrinal gap in Indian copyright law. While the Copyright Act 1957 and Information Technology Act 2000 provide the core rights and intermediary safe harbour respectively, neither statute explicitly contemplates algorithmic promotion of infringing content or the use of copyrighted material for AI training.[10] This creates uncertainty for platforms, creators, courts and regulators alike.
2. Tthe study situates India's challenges within comparative international practice. The European Union's Digital Single Market Directive (Article 17) imposes proactive obligations on platforms, whereas the United States' Digital

[8] Eastern Book Company & Ors v D. B. Modak & Anr AIR 2008 SC 809.
[9] Information Technology Act 2000 (India) s 79; ANI Media Pvt Ltd v OpenAI Inc & Anr CS(COMM) 1028/2024 (Delhi High Court).
[10] Copyright Act 1957 (India); Information Technology Act 2000 (India) s 79; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (India).

Millennium Copyright Act §512 remains reactive.[11] India has not yet chosen between these models. Analyzing their respective strengths and weaknesses offers policy makers a clear menu of options tailored to India's constitutional guarantees of free expression and due process.[12]

3. The study proposes calibrated reforms clarified notice standards, mandatory transparency, and statutory licensing for AI training that can be implemented within India's existing institutional architecture without chilling innovation or burdening small platforms. By directly tying the reforms to two core research issues intermediary liability for algorithmic promotion and regulation of AI training datasets this paper fills a gap in both legal scholarship and policy design.

More broadly, this research will help Indian creators, platforms and courts to navigate the complex trade-offs between effective enforcement and open digital innovation, thereby strengthening India's position in ongoing WIPO and WTO negotiations on AI and copyright.[13]

F. Research Questions

1. To what extent does Indian law clearly assign liability to intermediaries for algorithmic promotion or discovery of infringing content?

2. How should AI training datasets be regulated under Indian copyright law are current doctrines sufficient to require licences or permissions, or is legislative reform required?

G. Literature Review

1. Evolving Intermediary Liability in India

Indian scholarship has begun to re-examine the operation of intermediary safe harbour in the context of algorithmically curated platforms. In their article "Evolving

[11] Directive (EU) 2019/790 on Copyright in the Digital Single Market art 17; 17 U.S.C. § 512 (Digital Millennium Copyright Act).

[12] Constitution of India art 19(1)(a); Shreya Singhal v Union of India (2015) 5 SCC 1 (Supreme Court of India).

[13] WIPO, 'Guidelines on Text and Data Mining Exceptions' (2022); WTO, 'E-commerce and Intellectual Property: Ongoing Discussions' (2023).

Scope of Intermediary Liability in India", Indranath Gupta and Lakshmi Srinivasan observe that the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 have reignited debates concerning the scope of safe-harbour and the extent of obligations imposed on intermediaries. They demonstrate that the original statutory model envisaged intermediaries as passive conduits, but that this model no longer reflects the reality of large social-media and video-sharing platforms which rank, recommend and monetise content through proprietary algorithms.[14] Sriram Pradeep, in "Intermediary Liability: Recent Developments in India", similarly reviews post-2021 judicial decisions and concludes that Indian courts are moving incrementally towards requiring more proactive compliance from major platforms with takedown and grievance-redressal procedures even though the statutory text has not materially changed.[2] Both studies identify a latent doctrinal uncertainty highly relevant to this paper's first research question as to whether algorithmic promotion or recommendation of infringing content can itself be regarded as "knowledge" or inducement, thereby vitiating safe-harbour. This uncertainty underscores the absence of clear legislative or judicial guidance on the treatment of algorithmic conduct under Section 79 of the Information Technology Act.

2. Comparative Perspectives on Notice-and-Stay-Down

Internationally, Martin Husovec and co-authors' empirical study "The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown?" provides a systematic assessment of the advantages and risks of notice-and-takedown (NTD) and notice-and-stay-down (NSD) regimes. They demonstrate that NSD can significantly reduce repeat infringements but creates a structural risk of over-blocking lawful content unless accompanied by robust safeguards such as human review and transparent appeals. The Electronic Frontier Foundation's report "Unfiltered: How YouTube's Content ID Discourages Fair Use" supplies detailed case studies of automated filtering suppressing parody, criticism and other protected expression, illustrating the free-

[14] Indranath Gupta and Lakshmi Srinivasan, 'Evolving Scope of Intermediary Liability in India' (2023) https://pure.jgu.edu.in/5413/ accessed 20 September 2025.

expression costs of aggressive automated enforcement.[15] This comparative literature is pertinent to India's policy choice between NTD and NSD, yet no India-specific empirical study has evaluated the costs and benefits of such regimes calibrated to domestic platforms and creators a critical gap which this paper seeks to address.

3. AI Training Datasets and Copyright Risk in India

Legal commentary on the copyright status of text and data mining and the use of protected works in training machine-learning models is nascent in India. The Mondaq article "Training AI Models: Intersection Between AI and Copyright" provides one of the first detailed analyses of how the inclusion of copyrighted material in training datasets may constitute infringement under Section 51 of the Copyright Act, particularly for commercial developers. Annapurna Roy's ETtech article "AI model training on copyrighted data needs public review: Industry" records growing stakeholder pressure for amendments to the Copyright Act and clearer rules concerning fair-dealing exceptions and the licensing of datasets.[16] These commentaries reveal that, while the issue is under active industry and policy discussion, there is as yet no Indian statute or judicial decision that directly addresses whether large-scale ingestion of copyrighted works for machine-learning purposes is permissible or requires prior authorisation. The absence of any licensing mechanism or statutory clarification for AI-training datasets highlights a second major research gap that this paper will address.

IV. LEGAL FRAMEWORK AND JURISPRUDENCE ON INTERMEDIARY LIABILITY IN THE DIGITAL AGE

Indian copyright law is rooted in the Copyright Act 1957, which confers upon authors a bundle of exclusive rights over reproduction, adaptation and communication to the public, but conditions protection on the originality of the work. In *Eastern Book*

[15] Martin Husovec and others, 'The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown?' (SSRN, 2020) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3660625 accessed 20 September 2025.
[16] 'Training AI Models: Intersection Between AI and Copyright' (Mondaq, 2024) https://www.mondaq.com/india/copyright/1587932/ accessed 20 September 2025.

Company v D. B. Modak the Supreme Court rejected the older "sweat of the brow" approach and held that originality requires a minimal degree of creativity manifested through skill and judgement, even in compilations or editorial work.[17] This doctrinal shift is central because it defines the threshold for protection of works that may later be disseminated digitally and informs the claims that rights-holders may bring against intermediaries. Alongside the Copyright Act, the Information Technology Act 2000 introduces Section 79, which provides intermediaries with a conditional "safe harbour" from liability for third-party content, provided that they (a) do not initiate the transmission or modify the content, (b) observe due diligence and (c) act expeditiously upon receiving "actual knowledge" or a valid notice of infringement.[18] The Intermediary Guidelines Rules 2011 and the updated 2021 Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules supplement Section 79 by prescribing notice formats, takedown timelines and grievance redressal obligations, especially for "significant social media intermediaries." Together, these provisions reflect a legislative attempt to balance authors' rights and innovation by allocating direct liability to infringing users while shielding neutral intermediaries, but they remain rooted in a reactive, notice-driven paradigm developed for an earlier phase of the internet.

Judicial decisions illustrate both the operation and the limits of this regime. In *Eastern Book Company* the Court affirmed that editorial additions such as headnotes, paragraphing and cross-citations in law reports are protectable as original works, even though the underlying judgments are in the public domain.[19] This interpretation broadened the scope of copyright in India and underscores the stakes for digital intermediaries hosting large volumes of content. The Delhi High Court's decision in *MySpace Inc. v Super Cassettes Industries Ltd* is the leading Indian authority on intermediary liability for copyright infringement.[20] Super Cassettes alleged that

[17] Eastern Book Company & Ors v D. B. Modak & Anr AIR 2008 SC 809, paras 29–33.
[18] Information Technology Act 2000 (India) s 79; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (India).
[19] Eastern Book Company & Ors v D. B. Modak & Anr AIR 2008 SC 809.
[20] MySpace Inc v Super Cassettes Industries Ltd SCC OnLine Del 6382 (Delhi High Court, 23 December 2016).

MySpace hosted user-generated videos incorporating its music catalogue. The Court held that an intermediary does not lose safe harbour merely because infringement is possible or foreseeable, or because the platform has general awareness of infringing activity. Instead, liability turns on "actual knowledge" or specific notice of infringing material and a failure to act. The judgment also required complainants to provide exact details such as URLs so that intermediaries can act precisely, thereby preventing open-ended or vague takedown demands.[21] This jurisprudence clarifies that Indian law remains strongly notice-based and that intermediaries are not expected to monitor content proactively absent statutory obligation.

However, significant gaps remain when these doctrines are applied to algorithmically driven platforms. Neither Section 79 nor the 2021 Rules explicitly address whether recommender systems, ranking algorithms or other automated features that "promote" infringing content can give rise to constructive knowledge or inducement liability[7]. Indian courts have yet to decide whether an intermediary that monetises or boosts infringing content through algorithmic recommendation can still claim neutrality. Rights-holders also face evidentiary hurdles: without transparency obligations they may be unable to prove the causal link between algorithmic promotion and harm. Moreover, the current rules impose uniform takedown duties but do not differentiate between large platforms with sophisticated content-ID systems and small start-ups. These lacunae are particularly pressing given India's growing digital ecosystem and the rise of generative AI and machine-learning models, which depend on ingesting large corpora of data. Against this backdrop, the following chapters examine comparative approaches and develop calibrated reforms that clarify the scope of intermediary liability in the digital age.

V. COPYRIGHT AND AI-TRAINING DATASETS: COMPARATIVE AND EMERGING APPROACHES

In many jurisdictions, the advent of artificial intelligence and machine-learning has forced governments and courts to grapple with whether existing copyright exceptions

[21] R Ramanathan, 'Safe Harbour and Algorithmic Platforms: Rethinking Section 79 IT Act' (2022) NUJS L Rev.

(or new ones) should permit the use of protected works for training datasets. The European Union's Digital Single Market (DSM) Directive (2019) is among the clearest recent examples. Article 3 provides that research organizations and cultural heritage institutions may carry out text and data mining (TDM) for scientific research for works and other subject matter to which they lawfully have access, while Article 4 imposes a broader mandatory exception for TDM regardless of purpose, so long as the reproductions and extractions are lawfully made and rightsholders have not expressly reserved the rights, particularly via machine-readable means.[22] This setup grants both a baseline of legal certainty and a mechanism for balancing rightsholders' interests through the reservation opt-out provision. Compared with Europe, India currently lacks any specific statutory provision analogous to Articles 3 or 4 of the DSM Directive. Existing Indian law does not expressly address whether machine-learning training that ingests large volumes of copyrighted works is exempted under the narrow "fair dealing" exceptions, nor does it provide for a sui generis exception for TDM or AI-training.[23] The absence of such clarity places developers and rights-holders in a legal grey zone: whether the reuse is permissible depends heavily on case law, contractual terms, or tacit licensing arrangements, all of which are uncertain and uneven.

In the United States, the doctrine of fair use under Title 17 and the protections afforded by the DMCA provide a mix of flexibility and risk for AI training. U.S. fair use jurisprudence has yet to produce a binding, comprehensive decision that clearly states that building a foundational model using copyrighted texts is lawful per se; rather, disputes continue in class actions and complaints (for example, allegations against Meta's LLaMA model producers) that prompt consideration of fair use factors such as

[22] Legal basis of DSM Directive, especially Articles 3 and 4, providing exceptions and limitations for text and data mining, and the requirement of "machine-readable means" to reserve rights: see Legal Basis | TDM·AI (EU) https://docs.tdmai.org/legal-aspects/legal-basis accessed 20 September 2025; EU DSM Copyright Directive: Overview (Mondaq, UK) https://www.mondaq.com/uk/copyright/801438/eu-dsm-copyright-directive-overview accessed 20 September 2025.

[23] 'Text And Data Mining: Decoding Copyright Challenges In India' (India – Intellectual Property) (Mondaq, 2024) https://www.mondaq.com/india/copyright/1456378/text-and-data-mining-decoding-copyright-challenges-in-india accessed 20 September 2025.

purpose, amount used, and market substitution. Because fairness is judged case by case, outcomes are unpredictable, and rights-holders often assert that unauthorized training reproduces their works in ways that compete with or supplant legitimate markets. Meanwhile, developers argue that transformations inherent in model training, or the use of public domain or lawfully accessed materials, may weigh toward fair use. These tensions echo those in Europe, but without an opt-out reservation provision like that in DSM; instead, U.S. law leaves much to litigation and negotiation. Applying this model to India reveals that while fair dealing is the closest analogue, its narrow scope and lack of interpretive precedent for commercial or large-scale AI means it falls short of providing the legal certainty needed.

Germany has also begun clarifying its own regime in response to EU directives. A recent case in the Regional Court of Hamburg (Case 310 O 227/23) held that the creation of AI training datasets may not infringe copyright under German law when the copyrighted photograph used is lawfully accessible and when the use falls under a statutory text and data mining exception (§ 60d UrhG), which provides for a reserve mechanism and conditions for lawful access and storage.[24] This decision underscores how courts can interpret TDM exceptions expansively and how statutory text-based regulation (plus reservation clauses) supports permissible dataset creation. For India, such developments suggest that legislative drafting that includes mandatory exceptions for TDM (with reservation/opt-out rights for rightsholders), clarity on lawful access, and storage/retention conditions will be critical. Also, India's Digital Personal Data Protection Act (2023) and existing IP statutes currently do little to address the overlap between personal data use and copyrighted content in AI training, leaving legal uncertainty in privacy, IP and data protection overlap.[25]

[24] Hamburg Regional Court Judgment 310 O 227/23 – "Creation of AI Training Datasets Does Not Infringe Copyright", German UrhG § 60d, per §44b / §60d exceptions, as reported in SimonGraeser Law https://www.simongraeser.law/en/creation-of-ai-training-datasets/ accessed 20 September 2025; TDM Exceptions And Copyright: A German Court Decision (Mondaq Germany) https://www.mondaq.com/germany/copyright/1532854/tdm-exceptions-and-copyright-a-german-court-decision accessed 20 September 2025.

[25] 'Text And Data Mining: Decoding Copyright Challenges In India' (n 2); Bridge Counsels, 'Text and Data Mining vs. India's Digital Personal Data Protection Act, 2023: a Critical Study' https://bridgecounsels.com/text-and-data-mining-vs-indias-digital-personal-data-protection-act-

VI. RECOMMENDATIONS AND PROPOSED REFORMS

Based on the doctrinal and comparative analysis in the previous chapters, the following calibrated reforms are recommended for India's copyright and intermediary liability regime.

A. Clarify Intermediary Obligations

1. **Codify "actual knowledge" —** Amend Section 79 of the IT Act or include in the Digital India Act an express definition of "actual knowledge" to specify when algorithmic detection or recommendation of infringing content triggers a removal obligation.
2. **Standardize notice procedures —** Introduce a uniform statutory notice format with mandatory fields (work, URL/location, claimant's details, basis of claim) and fixed response timelines, coupled with a counter-notice mechanism to protect lawful uses and avoid over-blocking.

B. Mandatory Transparency and Algorithmic Disclosure

1. **Takedown transparency reports —** Require "significant social media intermediaries" to publish six-monthly statistics on the number of notices received, removal rates, appeals and reinstatements.
2. **Disclosure of recommender parameters —** Oblige large intermediaries to disclose, in general terms, the main parameters of recommender algorithms materially affecting the visibility of content flagged as infringing, without compelling disclosure of proprietary code.

C. Statutory or Collective Licensing for AI Training Datasets

1. **Extend compulsory licensing —** Amend Section 31D of the Copyright Act or enact a new provision authorising collective licences for machine-learning training datasets, with transparent royalty calculations and opt-out procedures for creators.

2023-a-critical-study-of-the-legal-gap-and-its-implications-for-ai-governance/ accessed 20 September 2025.

2. **Establish a Digital Copyright Licensing Authority —** Create a specialised body to administer such licences, audit dataset provenance and resolve disputes swiftly.

D. Pilot Hybrid Notice-and-Stay-Down for Large Platforms

1. **Limited stay-down obligation —** Implement a pilot hybrid system under which only very large platforms must adopt fingerprinting for verbatim matches combined with swift human review and an independent fast-appeals mechanism.
2. **Exemption for smaller intermediaries —** Maintain lighter obligations for small or start-up platforms to preserve innovation while enforcing rights on mass-reach platforms.

E. Capacity Building and International Cooperation

1. **Strengthen judicial and enforcement capacity —** Invest in digital forensic units and continuing judicial education on AI evidence, dataset licensing and algorithmic accountability.
2. **Negotiate international instruments —** Pursue bilateral MOUs for cross-border takedown/blocking requests and participate actively in WIPO/WTO processes to develop a model AI-training licensing framework aligned with international best practices.

VII. CONCLUSION

The analysis undertaken in this study demonstrates that the present Indian framework governing copyright and intermediary liability remains essentially reactive and fragmented. While the Copyright Act 1957 and the Information Technology Act 2000, supplemented by the 2021 Intermediary Guidelines, have created a foundation for protecting authors' rights and shielding neutral intermediaries, these provisions were conceived for an analogue environment and do not adequately address the systemic realities of algorithmically driven platforms or large-scale machine-learning practices. The jurisprudence developed by the Supreme Court in *Eastern Book Company v D. B. Modak* and by the Delhi High Court in *MySpace*

v Super Cassettes clarified originality and the contours of "actual knowledge" but left unresolved the position of platforms whose recommender systems actively shape the visibility and monetization of content.

The comparative survey of European, American and German approaches indicates that clearer statutory text-and-data-mining exceptions, opt-out mechanisms, and proportionate disclosure obligations can reconcile the legitimate interests of right-holders and the public interest in innovation. These models also illustrate the importance of collective licensing structures for activities such as AI training that rely on vast quantities of protected works and cannot be negotiated individually.

On the basis of these insights, this paper proposes a recalibration of Indian law that would move beyond a purely notice-based regime towards a balanced system combining defined knowledge standards, standardised notices and counter-notices, proportionate transparency duties for large platforms, statutory licensing for AI training datasets, and targeted capacity-building and international cooperation. Such measures would not replicate foreign regimes mechanically but would adapt their underlying principles to India's constitutional framework, safeguarding freedom of expression and due process while providing legal certainty and fair remuneration for creators.

A coherent and future-proof copyright regime is essential to maintain public trust in digital markets and to encourage both creativity and technological progress. By adopting clear, balanced and transparent rules, India has the opportunity to establish itself as a jurisdiction that protects authors without stifling innovation and that offers a model of equitable governance for the digital and artificial-intelligence era.

VIII. BIBLIOGRAPHY

A. Primary Legislation and Rules

1. Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), Marrakesh Agreement Establishing the World Trade Organization, Annex 1C (1994)

2. Berne Convention for the Protection of Literary and Artistic Works (Paris Act 1971)

3. Copyright Act 1957 (India)

4. Information Technology Act 2000 (India)

5. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (India)

6. Digital Personal Data Protection Act 2023 (India) (for overlap with AI training datasets)

B. Case Law

1. *Eastern Book Company and Others v D. B. Modak and Another* AIR 2008 SC 809 (Supreme Court of India)

2. *MySpace Inc v Super Cassettes Industries Ltd* SCC OnLine Del 6382 (Delhi High Court, 23 December 2016)

3. ANI Media Pvt Ltd v OpenAI Inc & Another CS(COMM) 1028/2024 (Delhi High Court)

4. Hamburg Regional Court Judgment 310 O 227/23 'Creation of AI Training Datasets Does Not Infringe Copyright' (Regional Court of Hamburg, Germany)

C. Books

1. Lionel Bently and Brad Sherman, *Intellectual Property Law* (6th edn, Oxford University Press 2022)

2. P Bernt Hugenholtz, *Collective Rights Management in the Digital Age* (JIPITEC 2019)

D. Journal Articles and Working Papers

1. Martin Husovec and others, 'The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown?' (2020) SSRN

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3660625 accessed 20 September 2025

2. R Ramanathan, 'Safe Harbour and Algorithmic Platforms: Rethinking Section 79 IT Act' (2022) NUJS Law Review

3. Bridge Counsels, 'Text and Data Mining vs India's Digital Personal Data Protection Act, 2023: a Critical Study of the Legal Gap and its Implications for AI Governance' (2024) https://bridgecounsels.com/text-and-data-mining-vs-indias-digital-personal-data-protection-act-2023-a-critical-study-of-the-legal-gap-and-its-implications-for-ai-governance/ accessed 20 September 2025

E. Reports, Guidelines and Blogs

1. OECD, 'Transparency and Accountability in Recommender Systems' (2023) https://www.oecd.org/digital/ accessed 20 September 2025

2. Electronic Frontier Foundation, 'Unfiltered: How YouTube's Content ID Discourages Fair Use' (2018) https://www.eff.org accessed 20 September 2025

'EU DSM Copyright Directive: Overview' (Mondaq, 2019) https://www.mondaq.com/uk/copyright/801438/eu-dsm-copyright-directive-overview accessed 20 September 2025

3. 'Text and Data Mining: Decoding Copyright Challenges in India' (Mondaq, 2024) https://www.mondaq.com/india/copyright/1456378/text-and-data-mining-decoding-copyright-challenges-in-india accessed 20 September 2025

4. Kluwer Copyright Blog, 'Generative AI: Admissibility and Infringement in the Two US Class Actions against Meta's LLaMA' (2024) https://copyrightblog.kluweriplaw.com/2024/01/17/generative-ai-admissibility-and-infringement-in-the-two-us-class-actions-against-metas-llama/ accessed 20 September 2025

5. SimonGraeser Law, 'Creation of AI Training Datasets Does Not Infringe Copyright' (2024) https://www.simongraeser.law/en/creation-of-ai-training-datasets/ accessed 20 September 2025

6. WIPO, 'Guidelines on Text and Data Mining Exceptions' (2022) https://www.wipo.int/ accessed 20 September 2025