Trends in Intellectual Property Research

Innovative

Article

Intellectual Property and Personal Data in AI Datasets Under India's DPDP Act 2023

Aman Kumar Jha*, Mohit Jain and Tanishk Bhawsar

National Law Institute University, Kerwa Dam Road, Bhopal-462044, India
Mohitjain.bscllb@nliu.ac.in (M.J); tanishkbhawsar.bscllb@nliu.ac.in (T.B)
* Correspondence: amankumarjha.bscllb@nliu.ac.in (A.K.J)

Abstract: The rapid expansion of generative AI has challenged India's fragmented legal regime governing AI training data, spanning the Copyright Act, 1957, trade secret protection, and the Digital Personal Data Protection Act, 2023 (DPDPA). This study explores the doctrinal incompatibility between India's purpose-specific fair dealing framework under Section 52 and the industrial-scale reproduction intrinsic to AI training, which fails the jurisdictional "purpose test" articulated in *Super Cassettes Industries Ltd v. Hamar Television Network* (2011). The study exposes the structural inadequacy of trade secret law in protecting the "composited value" of large-scale aggregated datasets, which lack the identifiability and durability required for conventional protection. The DPDPA's consent-centric architecture is functionally unworkable in billion-token training corpora characterized by attenuated data-principal relationships. Concrete doctrinal fault lines, including uncertainty surrounding "reproduction in material form" under Section 14(a)(i), transparency-trade secret conflicts identified in the DPIIT Working Paper on Generative AI and Copyright, and cross-border transfer constraints under Section 17 of the DPDPA have been mapped. Legal uncertainty will undermine both AI innovation and stakeholder protection in India if an integrated statutory framework for permissible training practices and rights allocation is not opted.

Keywords: AI training data; DPDPA compliance; copyright fair dealing; trade secrets; DPIIT framework; data ownership India

Citation: Aman Kumar Jha, Mohit Jain and Tanishk Bhawsar. 2026. Intellectual Property and Personal Data in AI Datasets Under India's DPDP Act 2023. *Trends in Intellectual Property Research* 4(2), 1-6. https://doi.org/10.69971/tipr.4.2.2026.109

Copyright: © 2026 by the authors. This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0.

1. Introduction

Generative artificial intelligence (GAI) systems have disrupted the traditional frameworks of data ownership and intellectual property protection (Kaplan and Samuel McCandlish 2020). Large language models (LLMs), diffusion models, and other sophisticated AI architectures require training on datasets of unprecedented scale and diversity. In India, this technological revolution has challenged three distinct legal regimes, namely copyright law, trade secret protection, and the newly operative Digital Personal Data Protection Act 2023 (DPDPA). These frameworks were not designed for AI-generated value chains, and their intersection undermines legal certainty for AI developers while inadequately protecting the rights of content creators and data principals (Buick 2025).

The question of data ownership in AI development is not merely academic. The recent Department for Promotion of Industry and Internal Trade (DPIIT) Working Paper on Generative AI and Copyright, India confronts a structural crisis in its approach to AI governance.[1] The current regime fails to establish clear ownership boundaries, leaving significant regulatory gaps, and creating perverse incentives for opacity rather than responsible innovation.

Indian law does not recognize an inherent "right to own data" in the manner it recognizes ownership of tangible property or copyright in literary works (Ugochukwu and Peter 2024). This absence reflects a deliberate policy choice. The Copyright Act 1957 protects creative expression, not raw information. The Indian Penal Code and other statutes protect trade secrets through confidentiality and breach of trust doctrines, not through ownership rights. The DPDPA 2023, despite

[1] Department for Promotion of Industry and Internal Trade. 2025.Working Paper on Generative AI and Copyright (Part I) Working Paper No 1. https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf

its sophisticated framework, grants data subjects rights of access and deletion rather than proprietary ownership.

This conceptual vacuum becomes acute in AI development. When an AI developer trains a system on a corpus of 500 billion text tokens drawn from public websites, academic databases, copyright-protected publications, and anonymized personal information, it's unclear who "owns" this aggregated training dataset. The developer who collected it, the individuals whose personal data it contains, the copyright holders whose works were reproduced or the platform operators who hosted the original sources? Indian law provides no coherent answer. The research gap here is not the absence of ownership theories, theoretical literature on data ownership is robust (Liu and L. Raymond Guo 2024).

Below dimensions define the core research gap. Indian law lacks not merely statutory provisions for AI training data, but fundamental doctrinal infrastructure capable of accommodating the nature of data aggregation, value creation, and information flows in AI development. Copyright law is designed to protect expressed ideas in fixed form, not the raw informational content extracted from those ideas (Wolff 2025). When an AI system ingests a million news articles to learn linguistic patterns, the copyright holder's exclusive right to reproduce the article under Section 14(a)(i) of the Copyright Act is technically infringed.[2] However, the infringement doctrine assumes human comprehension and consumption. The unauthorized copying contemplated by 19th century copyright law involved photocopying books, republishing newspapers, or mechanically reproducing musical recordings. The law treats the "reproduction" requirement as satisfied by acts that make the work accessible to human audiences.

Trade secret law protects specific, identifiable information conferring competitive advantage through confidentiality (Singh 2025). A customer list, a manufacturing process, a formula, these are discrete items capable of delimitation and protection. Training data for AI systems comprises billions of data points, each individually worthless but collectively valuable. A single article about renewable energy, extracted from a news database, has zero trade secret value. The same article as part of a 47 millionth token in a training dataset contributes incremental value that cannot be isolated or quantified.

DPDPA's framework centers on the consent and control of data subjects' individuals whose personal information is being processed.[3] This makes sense for personal data in traditional processing contexts. However, in AI training, the relationship between data subject and data user is attenuated almost beyond recognition. An individual who posted a medical question on a public forum in 2018 cannot realistically understand, assess, or control the downstream use of that data point as part of a training corpus in 2025 (Krimmelbein 2024).

2. Copyright Law and AI Training: The Failure of Fair Dealing

India's copyright framework diverges sharply from the flexible fair use doctrine of United States jurisprudence.[4] Section 52 of the Copyright Act 1957 enumerates specific purposes for which limited reproduction is permitted: criticism or review, private or personal use including research, and reporting of current affairs.[5] These categories are exhaustive, not illustrative. As the Delhi High Court established in *Super Cassettes Industries Ltd v. Hamar Television Network Pvt Ltd (2011)*, the enquiry operates in two sequential stages: first, whether the use falls within a statutory purpose; and whether the use was fair in manner and extent.[6] The threshold question is jurisdictional if the use does not satisfy the first gate, fairness becomes irrelevant.

This doctrinal architecture creates fundamental incompatibility with AI training practices. Industrial-scale machine learning for language models involves systematic reproduction of copyrighted works at scale that defies human-centric copyright concepts. These reproductions satisfy none of the statutory purposes. The AI developer is not criticizing or reviewing works or engaging with them for purposes of research in the traditional sense and is not reporting current affairs. The developer is engaging in something categorically different, i.e., automated feature extraction and statistical modelling.

The research gap here is the precise doctrinal incompleteness of the fair dealing framework. It's often cited that AI training "does not fall within fair dealing" without explaining the specific mechanism of this incompleteness. The analysis requires precision. The "research" purpose under Section 52(1)(a) has been interpreted narrowly in Indian jurisprudence. AI training is agnostic as to content meaning. The machine cannot "research" in the sense copyright jurisprudence has defined it.

The "criticism or review" purpose similarly fails. Criticism requires evaluation of the work's merits, i.e., reviewing requires commentary engaging with the work's content. When an AI system reproduces a copyrighted novel to extract linguistic features, it is neither criticizing nor reviewing. It is treating the novel as raw material for computational purposes entirely divorced from the work's expressive content (Priyadarshi 2026). The second gap concerns whether AI training constitutes "reproduction" at all, and if so, at what point. Section 14(a)(i) grants copyright owners the exclusive right to reproduce the work "in any material form including storing."[7] The phrase "storing" appears to encompass storage on servers. However, doctrinal uncertainty persists regarding the purpose and function of stored copies.

In traditional copyright scenarios, storage is instrumentally connected to human access. In AI training, storage serves an entirely different function. The dataset is stored not for human access but for algorithmic processing. Moreover, the work is not stored in recognizable form and is decomposed into tokens, embedded as vectors, and transformed beyond human readability (Anonymous 2023). Whether this transformation satisfies the "material form" requirement is judicially unexplored in India. The Delhi High Court, in the *Asian News International v. OpenAI* (2024), is currently examining this precise question but has not yet ruled.[8] No

---

2 The Copyright Act, 1957 (14 of 1957) s 14(a)(i).

3 The Digital Personal Data Protection Act, 2023 (22 of 2023) s 7(2).

4 Buick (n 2).

5 The Copyright Act, 1957 (14 of 1957) s 52.

6 *Super Cassettes Industries Ltd v Hamar Television Network Pvt Ltd* 2011 (1) SCC 534.

7 The Copyright Act, 1957 (14 of 1957) s 14(a)(i).

8 *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr.*, CS(COMM) 1028/2024 (Del. HC).

Indian appellate court has definitively established whether ingestion of copyrighted content into neural network architectures constitutes reproduction.

The European Union, Japan, and several other jurisdictions have enacted explicit text-and-data-mining (TDM) exceptions to copyright law (OECD 2025). The EU Copyright Directive 2019 permits automated analysis of copyrighted works for research purposes.[9] India's statutory framework operates in the opposite direction. Any reproduction not falling within enumerated fair dealing purposes is presumptive infringement, with the burden on the user to establish permissibility. Reversing this default through statutory exception requires doctrinal reconceptualization of the relationship between copyright owners and technology developers (Kemp 2020).

Indian copyright law provides for seizure of infringing goods, conversion damages, and accounts of profits in manner more severe than European approaches.[10] These remedial frameworks were designed for tangible piracy photocopying, selling counterfeit books, manufacturing bootleg recordings. Applying these remedies to AI training would be both overinclusive and inadequate. The absence of a flexible injunctive framework comparable to US equity jurisprudence means courts lack appropriate tools for balancing developer and rights holder interests (Kupferschmid 2024).

3. Trade Secret Law: Protection without Ownership

Trade secrets occupy a peculiar niche in Indian intellectual property laws. Unlike copyright or patents, which create positive rights of ownership, trade secrets rely on negative protection, i.e., the law prohibits misappropriation through breach of confidence or violation of legal duty. This distinction becomes critical in the AI context. The Trade Secrets Protection Scheme emerged through common law doctrines of breach of confidence, later codified in statutory amendments. Protection requires three elements, namely the information is not publicly known, it confers competitive advantage through secrecy and the owner has taken reasonable measures to maintain confidentiality.

Applying this framework to AI training data reveals fundamental mismatch. Training datasets often comprise publicly available information like news articles, scientific papers, web content, open-source code. Once the information is publicly available, it ceases to satisfy the first element of trade secret protection (Mathur and Ananya 2025). The doctrine contemplates situations where an employee discloses a customer list or manufacturing formula. It does not contemplate aggregation of millions of publicly available documents. The corpus, as an aggregated collection, might possess trade secret status through investment and curation involved in creating it. However, this protection is fragile: once any element becomes publicly known through disclosure or reverse engineering, the trade secret status of the entire dataset collapses. The research gap concerns inadequacy of trade secret law for protecting "composited value" value arising from aggregation and curation of information none of which is individually secret (Wolff 2025). The EU attempted to address this gap through database rights, which protect the investment in collection and arrangement separately from copyright.[11] India's Copyright Act recognizes no comparable protection.

A more acute research gap emerges from conflict between trade secret protection and transparency obligations imposed by DPDPA and emerging copyright frameworks. The DPIIT Working Paper recommends that AI developers disclose summaries of the copyrighted content used in training datasets.[12] This is incompatible with trade secret protection. If the training corpus is deemed a trade secret, then requiring disclosure violates the foundational principle of secrecy. Indian law has no mechanism for balancing these competing interests.[13] The DPDPA Section 6(2) permits processing without consent where required by law, but provides no framework for balancing data principals' rights against developers' legitimate business interests.[14] The result is a binary choice of either developers maintain opacity about training data sources (violating emerging copyright transparency norms) or they disclose and lose trade.

4. DPDPA Obligations and the Attenuation of Consent

4.1 The Consent Framework and its Limitations

The DPDPA 2023 is India's comprehensive response to personal data protection concerns.[15] Unlike earlier sectoral approaches, the DPDPA applies across all sectors and establishes consent as the foundational principle governing personal data processing.[16] Section 7(2) requires that processing must be based on one of four grounds, namely consent, contract performance, legal obligation, or vital interests.[17] The research gap concerning DPDPA consent requirements relates to fundamental mismatch between the consent paradigm and realities of AI training. Consent doctrine presupposes that data subjects can understand the purpose for which their personal data will be used, assess the risks and benefits associated with such use, exercise meaningful choice regarding disclosure and withdraw consent if circumstances change. None of these premises hold in AI training scenarios. For example, an individual post a medical question on an online forum in 2018. A decade later, this data point becomes part of a training corpus for

---

9 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 and 4.

10 The Copyright Act, 1957 (14 of 1957) ss 55,56,57,58,59,60,61,62,63,64 and 65.

11 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

12 Department for Promotion of Industry and Internal Trade (n 3).

13 The Digital Personal Data Protection Act, 2023 (22 of 2023) s 6(2).

14 ibid.

15 The Digital Personal Data Protection Act, 2023 (22 of 2023).

16 The Digital Personal Data Protection Act, 2023 (22 of 2023) s7.

17 ibid.

an AI system that generates medical advice. The individual did not consent to use of their data for AI training and no meaningful mechanism exists for the individual to withdraw consent retroactively (Anonymous 2025).

Retrofitting consent doctrine to accommodate this scenario requires *ex-post facto* consent, where individuals are informed after inclusion in training datasets and asked to consent or withdraw. This is operationally infeasible at scale. In constructive consent, posting data on a public platform is deemed implicit consent to downstream uses. This inverts DPDPA's foundational principle that consent must be explicit, informed, and capable of withdrawal. While in consent waiver, individual consent is replaced with collective governance mechanisms. This is not contemplated in DPDPA as currently drafted. The Act provides limited recognition of these difficulties. Section 8 permits processing without consent where mandated by law, required for contract performance, or necessary for vital interests.[18] This provision is narrow and does not comprehensively address AI training scenarios.

4.2 Regulatory Uncertainty and Institutional Gaps

A more structural research gap is the institutional framework for DPDPA enforcement in the AI context. The DPDPA establishes the Data Protection Board, a regulatory body tasked with investigating complaints and enforcing compliance.[19] However, the Board's composition, decision-making authority, and appeal procedures remain substantially undefined in the primary legislation.[20] This institutional vagueness creates uncertainty particularly acute in cross-cutting issues like AI training data. The DPDPA lacks clarity on how its obligations interact with other regulatory regimes. Does the exemption for processing "where required by law" extend to processing required for copyright compliance or trade secret protection? If an AI developer discloses the provenance of training data to satisfy copyright transparency obligations, does this disclosure violate DPDPA's confidentiality requirements? Indian law provides no answer (Sarthak 2025). The DPDPA's provision for "sensitive personal data" exacerbates this uncertainty. Health, financial, and biometric categories receive heightened protection, requiring explicit consent even for processing.[21] AI systems trained on medical literature, financial reports, or biometric databases would necessarily process sensitive data. Yet the Act has no mechanism for collective consent or regulatory approval that would enable large-scale training while respecting enhanced protections.

4.3 The Cross-Border Data Transfer Dilemma

DPDPA Section 17 restricts cross-border transfer of personal data to designated countries or with explicit consent.[22] This creates distinctive Indian dimension to the training data ownership question. Many AI developers are based outside India. If personal data derived from Indian data subjects is transferred to servers outside India for AI training, DPDPA restrictions apply. Yet the Act provides no safe harbor for technological necessity or for transfers to companies with comparable data protection standards (Latham and Watkins 2023). The result is a potential block on India's participation in global AI development chains unless developers either obtain explicit consent from millions of Indian data subjects, or maintain separate training infrastructure within India. This gap is not merely regulatory but economic. Smaller AI companies cannot maintain geographically compartmentalized infrastructure.

5. The DPIIT Framework: Partial Solutions and Persistent Gaps

The DPIIT Working Paper (December 2025) represents the first comprehensive Indian policy engagement with AI and copyright.[23] The Committee recommends a hybrid licensing model wherein AI developers receive a blanket license to use lawfully accessed copyrighted content for training purposes, with royalties becoming due only upon commercialization of AI systems, at rates determined by a government-appointed body (Anonymous 2024). This framework attempts to reconcile copyright holders' rights with developers' practical needs. However, the DPIIT proposal, while innovative, does not address the ownership question directly and leaves critical gaps unresolved. The proposal applies only to copyright-protected content. It establishes no framework for determining ownership of aggregated datasets, allocating rights among multiple contributors, or protecting the investment in data curation.[24] An AI developer who invests substantially in identifying, accessing, and organizing training data receives no ownership protection beyond trade secret law's fragile confidentiality doctrine.

The proposal creates a false equivalence between different forms of content. The licensing framework assumes all copyrighted content has identifiable rights holders capable of participating in collective licensing arrangements. However, much copyrighted content is orphaned works whose copyright holders cannot be identified or located.[25] The proposal does not address how orphaned works should be treated. The proposal fails to address personal data.[26] The DPIIT Committee's mandate explicitly excluded DPDPA compliance questions, treating copyright and data protection as separate regimes. This is problematic. Training data typically comprises both copyright-protected content and personal data. The licensing framework addresses only the copyright dimension; personal data questions remain unresolved. The proposal does not address trade secrets. While the DPIIT recognizes that transparency requirements might compromise developers' trade secrets, it offers no mechanism for reconciling transparency obligations with trade secret protection.[27] This remains vague and operationally insufficient.

---

18 The Digital Personal Data Protection Act, 2023 (22 of 2023) s 8.

19 The Digital Personal Data Protection Act, 2023 (22 of 2023) ss 18,19,20,21,22& 23.

20 The Digital Personal Data Protection Rules, 2025 S.O. 846(E).

21 The Digital Personal Data Protection Act, 2023 (22 of 2023) s 10.

22 The Digital Personal Data Protection Act, 2023 (22 of 2023) s 17.

23 Department for Promotion of Industry and Internal Trade (n 3).

24 Department for Promotion of Industry and Internal Trade (n 3).c

25 ibid.

26 ibid.

27 ibid.

6. Comparative Insights and Why Importation is Insufficient

The EU approach comprises three distinct layers of the Copyright Directive (2019) with TDM exceptions; the AI Act (2024) with transparency requirements; and GDPR with enhanced protections for automated decision-making.[28] This layered approach presupposes a regulatory environment where copyright, data protection, and competition law have evolved in mutual awareness and coordination. India's legal framework emerged independently through separate statutes with minimal coordination. The EU's approach reflects philosophical commitment to treating data as strategic public resource requiring governance for broader innovation objectives.[29] India's DPDPA reflects more individualistic philosophy as personal data is a resource belonging to the individual, subject to that individual's control. These foundational premises are incompatible; one cannot graft EU's public-resource framework onto DPDPA's individual-control framework without doctrinal confusion (Anonymous 2021).

United States jurisprudence has is flexible and accommodates technological change through adaptive interpretation of copyright's fair use doctrine.[30] The Second Circuit's decision in *Authors Guild v. Google* (2015) established that Google's scanning of millions of books for search indexing constituted transformative fair use despite large-scale reproduction without permission.[31] However, Indian courts have never adopted the US approach's flexibility.[32] Indian jurisprudence treats fair dealing purposes as exhaustive rather than illustrative, and Indian courts have been reluctant to expand fair dealing beyond enumerated purposes.

The US approach relies on robust statutory safe harbors, particularly the Digital Millennium Copyright Act's Section 512 provisions, which shield online service providers from liability for user-generated infringement.[33] India has no comparable framework. The Copyright Act contains no safe harbor for technology platforms. Some jurisdictions have begun experimenting with explicit data ownership frameworks (Sengupta 2024). Singapore's Personal Data Protection Act (2012) contemplates compensatory mechanisms where data is used for research purposes. [34] The EU's proposed Data Act (2023/2854) establishes data access rights and contemplates sharing mechanisms where data has been generated through IoT devices.[35] These frameworks suggest that the global trend involves recognizing explicit data ownership subject to sharing obligations, rather than treating data as ownerless information.

7. Research Gaps and Implications

Synthesizing the foregoing analysis, several research gaps emerge that Indian scholarship has not adequately addressed. Indian copyright, trade secret, and data protection law are developed independently with no consideration of mutual interaction in AI contexts (Chimni and Vrinda Patodia 2024). The result is not merely gaps but contradictions, i.e., copyright law prescribes transparency regarding training data, trade secret law prescribes opacity and DPDPA prescribes individual control. These regimes cannot coexist without hierarchical resolution that Indian law has not established. When personal data, copyright-protected content, and developer investment all contribute to a training dataset's value, existing law provides no mechanism for allocating ownership or determining respective rights (Liu and L. Raymond Guo 2024). This gap becomes critical in multi-stakeholder scenarios. Indian copyright law's enforcement mechanisms (seizure, destruction, damages) are calibrated for tangible piracy. They are both overinclusive and inadequate for AI training contexts.[36] Courts require more nuanced remedial tools.

DPDPA's consent framework presupposes individual agency and meaningful choice. In large-scale data aggregation, these presuppositions fail.[37] Indian law provides no framework for collective consent, representative decision-making, or regulatory approval substituting for individual consent. AI development is inherently global (Shekar 2025). A training dataset might comprise content and personal data from dozens of jurisdictions, each with distinct copyright, data protection, and trade secret regimes. Indian law provides no framework for navigating this jurisdictional complexity. The fragmented legal landscape creates substantial uncertainty regarding the legality of AI training practices (Gupta and Nancy Roy 2025). This uncertainty itself is a barrier to innovation, particularly for smaller developers who cannot afford extended legal analysis.

8. Conclusions

"Who owns training data" reveals profound gaps in Indian law's conceptual, doctrinal, and institutional infrastructure for AI governance. The problem is not that courts or regulators have failed to resolve specific disputes. In fact, the relevant statutes, case law, and regulatory frameworks were not designed for scenarios involving large-scale data aggregation, multi-stakeholder value creation, and technological transformation beyond human comprehension. Retrofitting 19th century copyright doctrine, confidentiality-based trade secret protection, and individually-centred data protection to accommodate 21st century AI development requires not incremental amendment but structural legal reconstruction. The DPIIT's initiative toward hybrid licensing represents a positive first step. However, it addresses only copyright and ignores personal data dimensions, trade secret tensions, and ownership alloca-

---

28 ibid.

29 European Commission, *A European strategy for data* COM(2020) 66 final.

30 *Authors Guild v. Google, Inc.* 804 F.3d 202 (2d Cir. 2015).

31 ibid.

32 Priyadarshi (n 16).

33 Digital Millennium Copyright Act 1998 17 USC § 512 (US).

34 Personal Data Protection Act 2012 (Singapore)

35 Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (Data Act) [2023] OJ L 2023/2854

36 The Copyright Act, 1957 (14 of 1957) ss 55,56,57,58,59,60,61,62,63,64 and 65.

37 The Digital Personal Data Protection Act, 2023 (22 of 2023) s 7.

tion. Comprehensive reform requires statutory amendment establishing explicit frameworks for permissible AI training within defined parameters; allocation of rights among copyright holders, data subjects, and developers; remedial mechanisms appropriate to AI contexts; consent operation at scale; and international coordination. Until such comprehensive reform occurs, substantial legal uncertainty will persist, potentially inhibiting India's participation in global AI innovation while inadequately protecting the rights of content creators and data subjects whose information trains these systems.

References

Anonymous. 2021. Invoking trade secrets to block a request to access personal data under the GDPR: A "Threat" Has to Be Clearly Demonstrated. https://www.crowelltradesecretstrends.com/2021/04/invoking-trade-secrets-to-block-a-request-to-access-personal-data-under-the-gdpr/

Anonymous. 2023. AI and the Law of Copyright in India. https://spiceroutelegal.com/publications/ai-and-the-law-of-copyright/3/

Anonymous. 2024. Exploring the DPIIT's Working Paper on Generative AI and Copyright. https://www.ikigailaw.com/article/655/exploring-the-dpiits-working-paper-on-generative-ai-and-copyrigh

Anonymous. 2025. How Will the DPDPA Impact AI? https://www.dpdpconsultants.com/blog.php?id=38&title=how-will-the-dpdpa-impact-ai

Buick, Adam. 2025. Copyright and AI Training Data—Transparency to the Rescue? *Journal of Intellectual Property Law & Practice* 20: 182-192. https://doi.org/10.1093/jiplp/jpae102

Chimni, Arzu, and Vrinda Patodia. 2024. The DPIIT Working Paper on AI and Copyright: Regulatory signals and practical implications.https://www.obhanandassociates.com/blog/the-dpiit-working-paper-on-ai-and-copyright-regulatory-signals-and-practical-implications/

Gupta, Gaurav, and Nancy Roy. 2025. Text and data mining vs India's Digital Personal Data Protection Act, 2023: a critical study of the legal landscape.https://bridgecounsels.com/text-and-data-mining-vs-indias-digital-personal-data-protection-act-2023-a-critical-study-of-the-lega/

Sarthak, K. 2025. Copyright protection in LLM AI training – Part 2. https://www.khuranaandkhurana.com/2025/01/27/copyright-protection-in-llm-ai-training-part-2/

Kaplan, Jared, and Samuel McCandlish. 2020. Scaling laws for neural language models. *Journal of Machine Learning Research* 21: 1–30. https://arxiv.org/abs/2001.08361

Kemp, Richard. 2020. Algo IP: Intellectual property in AI datasets, insights and outputs – the growing importance of trade secrets. *Kemp IT Law*. https://kempitlaw.com/insights/algo-ip-intellectual-property-in-ai-datasets-insights-and-outputs-the-growing-importance-of-trade

Krimmelbein, Fred. 2024. Data ownership in the age of AI: the impact of data governance. https://labs.sogeti.com/data-ownership-in-the-age-of-ai-the-impact-of-data-governance/

Kupferschmid, Keith. 2024. Requiring AI transparency won't destroy the trade secrets of AI Companies. https://copyrightalliance.org/ai-transparency/

Latham and Watkins. 2023. India's digital personal data protection Act 2023 vs. the GDPR: a comparison. https://www.lw.com/admin/upload/SiteAttachments/Indias-Digital-Personal-Data-Protection-Act-2023-vs-the-GDPR-A-Comparison.pdf

Liu, Shuimei, and L. Raymond Guo. 2024. Data Ownership in the AI-Powered Integrative Health Care Landscape. *JMIR Medical Informatics* 12. https://doi.org/10.2196/57754

Mathur, Arnav, and Ananya Popli. 2024. Trade, privacy and DPDPA: crafting India's response to the privacy-trade dilemma. https://nliulawreview.nliu.ac.in/blog/trade-privacy-and-dpdpa-crafting-indias-response-to-the-privacy-trade-dilemma/

Organization for Economic Co-operation and Development. 2025. Intellectual property issues in artificial intelligence trained on scraped data. https://www.oecd.org/en/publications/intellectual-property-issues-in-artificial -intelligence-trained-on-scraped-data_d5241a23-en.html

Priyadarshi, Shubhi. 2026. Fair dealing cannot be presumed: why AI Training fails the purpose test under indian copyright law. https://www.barandbench.com/columns/fair-dealing-cannot-be-presumed-why-ai-training-fails-the-purpose-test-under-indian-copyright-law

Sengupta, Pamela. 2024. Data ownership and privacy in the age of generative AI. https://www.ve3.global/data-ownership-and-privacy-in-the-age-of-generative-ai/

Shekar, Shaurya. 2025. Training AI, testing law: India's copyright challenge with TDM. https://lawschoolpolicyreview.com/2025/08/08/training-ai-testing-law-indias-copyright-challenge-with-tdm/

Singh, Satyam. 2025. Double-edged data: the trade secret dilemma in India's DPDP act. https://www.mondaq.com/india/trade-secrets/1643096/double-edged-data-the-trade-secret-dilemma-in-indias-dpdp-act

Ugochukwu, Albert I., and Peter W. B. Phillips. 2024. Open data ownership and sharing: challenges and opportunities for application of FAIR principles and a checklist for data managers. *Journal of Agriculture and Food Research* 16: 1-9. https://doi.org/10.1016/j.jafr.2024.101157

Wolff, Yves-Alexander. 2025. AI training data sets as intellectual property? protectability and protection gaps of training data and data sets. https://buse.de/en/blog-en/technology/ai-training-data/