The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2026

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

TRADE SECRETS v. ALGORITHMIC TRANSPARENCY IN AI: HOW INDIAN DATA PROTECTION AUTHORITIES CAN AUDIT PROPRIETARY LLMs WITHOUT VIOLATING IP

~ Ms. Aashi Lalchand Khanchandani

ABSTRACT

The rapid proliferation of Large Language Models in India's financial, healthcare, and public-service sectors has intensified a fundamental tension in information law: the right of AI developers to protect commercially sensitive model designs and training data as trade secrets, versus the public interest in auditing these systems for bias, privacy infringements, and accountability failures. The Digital Personal Data Protection Act, 2023 (hereinafter referred to as DPDP Act), together with the DPDP Rules, 2025, establishes an initial yet incomplete framework for algorithmic oversight. However, the absence of a dedicated trade secrets statute leaves intellectual property protection reliant on common law, contract, and equity. This article analyzes the doctrinal divisions between these competing imperatives, reviews international audit models, particularly the EU AI Act, 2024, and proposes a structured, IP-respecting audit architecture for Indian data protection authorities. The proposed framework incorporates black-box testing methodologies, tiered secrecy protocols, and regulatory sandboxes. It is argued that the perceived binary between disclosure and secrecy is misleading; robust audit frameworks can achieve regulatory accountability without necessitating disclosure of proprietary model weights, architectures, or training corpora.

Keywords *: Large Language Models, Algorithmic Transparency, Trade Secrets, DPDP Act 2023, EU AI Act.*

I. INTRODUCTION

Large Language Models *(hereinafter referred to as "LLMs")* are no longer confined to experimental settings. They are currently integrated into credit-scoring engines, judicial case-management systems, medical diagnostic platforms, and public-grievance portals throughout India. The India AI Mission aims to position the country as a global center for AI innovation, with the domestic AI market projected to reach USD 17 billion by 2027.[1] However, the architecture that underpins the effectiveness of LLMs billions of parameters trained on extensive, frequently opaque datasets makes them inherently resistant to traditional regulatory inspection.

This opacity gives rise to two distinct yet interconnected concerns. Firstly, the regulators and civil society, the *"black-box"* nature of LLMs raises issues related to encoded bias, privacy violations resulting from memorized training data, and a lack of meaningful accountability when automated decisions result in harm. Secondly, for the AI developers, model weights, fine-tuning datasets, and architectural innovations that differentiate one LLM from another constitute critical intellectual property assets, whose competitive value relies on secrecy. Developers contend that compelled disclosure to regulators is effectively equivalent to disclosure to competitors.

The Indian law occupies a position at the intersection of these concerns but does not yet offer a coherent solution. The DPDP Act, 2023, designates certain high-risk entities as Significant Data Fiduciaries (*hereinafter referred to as "SDFs"*) and requires independent data audits and Data Protection Impact Assessments *(hereinafter referred to as "DPIAs")*, yet it remains largely silent on the manner in which auditors should interact with proprietary algorithmic systems.[2] India lacks a dedicated trade secrets statute, although a Trade Secrets Bill is under legislative consideration; currently, protection is afforded through contract law, equitable principles, and the Indian Contract Act, 1872.[3]

---

[1] *AI Regulations in India 2025: Complete Compliance Guide*, American Chase (Dec. 11, 2025), https://americanchase.com/generative-ai-regulations-india/ (last accessed on June 14, 2026)

[2] Digital Personal Data Protection Act, 2023, No. 22 of 2023, § 10 (India); Digital Personal Data Protection Rules, 2025, Rule 13.

[3] *An Indian Trade Secret Bill May Reshape the AI x Copyright Dead-end: An Explainer*, The Legal Wire (Nov. 10, 2025), https://thelegalwire.ai/an-indian-trade-secret-bill-may-reshape-the-ai-x-copyright-dead-end-an-explainer/ (last accessed on June 14, 2026)

II. TRADE SECRETS AND LLMs IN INDIA: THE EXISTING LEGAL FRAMEWORK

A. Absence of a Dedicated Statute

In contrast to the United States, which codified trade secret protection through the Defend Trade Secrets Act, 2016, and the European Union, which enacted Directive 2016/943 on the protection of undisclosed know-how, India does not have a dedicated trade secrets statute.[4] Instead, courts enforce confidentiality obligations through contract law, equity, and tort principles.[5] Non-disclosure agreements, restrictive covenants, and confidentiality clauses in employment contracts serve as the primary mechanisms for protection. Indian courts have shown a willingness to grant injunctive relief when contractual safeguards are clearly established. For instance, in ***Bombay Dyeing and Manufacturing Co. Ltd. v. Mehar Karan Singh***[6], the court restrained a former employee from misusing proprietary algorithmic data and customer information where a valid NDA was in place.

The absence of dedicated legislation has attracted significant attention. As of 2025, a Trade Secrets Bill is under active consideration in India. According to The Legal Wire, proposed provisions under Section 3 of the draft Bill would protect any information *"deriving commercial value on account of being secret"* and subject to reasonable steps to maintain secrecy. This formulation closely aligns with Article 39 of the TRIPS Agreement.[7] As for the AI developers, such provisions could enable model weights, training data, and internal tooling to be formally recognized as protectable trade secrets, independent of any contractual arrangement.

B. LLMs as Trade Secrets: Scope and Limitations

The core LLM architectures, loss functions, and training methodologies receive limited patent protection in India. Section 3(k) of the Patents Act, 1970, excludes *"a computer programme per se"* from patentability. Although, the Draft Computer-Related Inventions (CRI) Guidelines,

---

[4] Monika Bartczak, *The Uniformization of the Principles of Trade Secret Protection in the European Union Law, The Implications for Polish Law: A Study of Selected Issues*, 1(10) TORUN INT'L STUD. 139, 139 (2018), https://doi.org/10.12775/tsm.2017.011 (last accessed on June 14, 2026)

[5] *Patent & Trade-Secret Playbooks for Software and AI Models in India*, Mondaq (Sept. 4, 2025), https://www.mondaq.com/india/trade-secrets/1673654/patent-trade-secret-playbooks-for-software-and-ai-models-in-india. (last accessed on June 14, 2026)

[6] Bombay Dyeing & Mfg. Co. Ltd. v. Mehar Karan Singh, (2019) (Delhi H.C.).

[7] *An Indian Trade Secret Bill May Reshape the AI x Copyright Dead-end: An Explainer*, *supra* note 3; Agreement on Trade-Related Aspects of Intellectual Property Rights art. 39, Apr. 15, 1994, 1869 U.N.T.S. 299.

March 2025, issued by the Indian Patent Office, recognize that Artificial Intelligence and Machine Learning (AI/ML) models demonstrating a concrete technical contribution may qualify for protection, core algorithmic innovations such as transformer architectures, attention mechanisms, and reinforcement learning from human feedback (RLHF) protocols remain largely unpatentable.[8] Consequently, trade secrecy has become the predominant intellectual property strategy for LLM inventors in India.

LLM trade secrets encompass several distinct components: (i) model weights - which are billions of numerical parameters encoding learned relationships; (ii) training data composition - including data sources, curation methodologies, and the proportional mix of data used for pre-training and fine-tuning; (iii) architectural choices - such as layer configurations, context window optimizations, and proprietary inference enhancements; and (iv) evaluation benchmarks and red-teaming data - which are internal safety assessments revealing both capabilities and vulnerabilities. The Ocean Tomo framework identifies each of these elements as possessing *"commercial value on account of being secret."*[9] Requiring disclosure of these components to a regulator, even under confidentiality obligations, introduces risks of inadvertent leakage, cyber-theft from regulatory systems, and reverse engineering from partial disclosures.

III. THE DPDP FOUNDATION: AUDIT OBLIGATIONS AND IDENTIFIED GAPS

A. Significant Data Fiduciaries and the Audit Requirements

The DPDP Act, 2023 establishes a two-tier compliance structure. All Data Fiduciaries, defined as entities determining the purpose and means of processing personal data, are subject to general obligations under Section 8. These obligations include data minimization, accuracy, implementation of security safeguards, and breach notification. A subset of these entities, designated by the Central Government as SDFs under Section 10, are subject to additional requirements. SDFs must appoint a Data Protection Officer (DPO) resident in India, engage an

[8] Indian Patent Office, Draft Guidelines for Examination of Computer Related Inventions (Mar. 2025); The Patents Act, 1970, No. 39 of 1970, § 3(k).
[9] *Trade Secret Protection in the Age of Large Language Models: Risks, Reasonable Measures, and Legal Remedies*, Ocean Tomo (June 25, 2025), https://oceantomo.com/insights/trade-secret-protection-in-the-age-of-large-language-models-risks-reasonable-measures-and-legal-remedies/ (last accessed on June 14, 2026)

independent data auditor to assess compliance with the Act, and conduct periodic DPIAs along with other prescribed measures.[10]

The DPDP Rules, 2025 implement these statutory requirements. Rule 13 requires SDFs to conduct annual DPIAs and audits beginning from the date of their designation. Significant findings and identified gaps must be reported periodically to the Data Protection Board of India (DPBI).[11] The Data Protection Board, which serves as the Act's investigatory and adjudicatory authority, is vested with powers equivalent to those of a civil court under Section 28(7). These powers include summoning individuals, receiving evidence, and inspecting documents.[12]

B. The Algorithmic Blind Spot

Despite this framework, the DPDP Act remains largely silent on auditor' engagement with proprietary algorithmic systems. The audit mandate focuses on compliance with the Act's data protection provisions, such as consent, purpose limitation, and security measures, rather than addressing algorithmic fairness or explainability. The De Facto Law Journal notes that India's data protection *law does not mandate audits, explainability, or transparency for AI systems used in critical sectors like healthcare, law enforcement, and financial services and further observes that the complexities of AI decision-making like data inference or algorithmic bias are not adequately covered.*[13]

This gap holds constitutional significance. The Supreme Court's landmark decision in ***Justice K.S. Puttaswamy (Retd.) v. Union of India,***[14] established privacy as a fundamental right under Articles 14, 19, and 21 of the Constitution. Subsequent scholarship has expanded the Puttaswamy framework to include informational self-determination, which encompasses the right of individuals to understand and contest automated decisions affecting them. When a LLM is used

[10] DPDP Act, *supra* note 2, § 10(2)(a)–(c).
[11] DPDP Rules, *supra* note 2, Rule 13; *Obligations of Significant Data Fiduciaries Under the DPDP Act, 2023 and DPDP Rules, 2025*, Tsaaro (Nov. 15, 2025), https://tsaaro.com/blogs/obligations-of-significant-data-fiduciaries-under-the-dpdp-act-2023-and-dpdp-rules-2025/ (last accessed on June 14, 2026).
[12] DPDP Act, *supra* note 2, § 28(7); *The Digital Personal Data Protection Act of India, Explained*, Future of Privacy Forum, https://fpf.org/blog/the-digital-personal-data-protection-act-of-india-explained/ (last accessed on June 14, 2026).
[13] *Why We Need Data Protection Laws for AI in India*, De Facto L.J. (May 12, 2025), https://defactolawjournal.org/papers/why-we-need-data-protection-laws-for-ai-in-india/ (last accessed on June 14, 2026).
[14] Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 S.C.C. 1

to make decisions that *"affect the Data Principal"*, Section 8(3) of the DPDP Act obligates the processing entity to ensure the completeness, accuracy, and consistency of the personal data used. This requirement implicitly acknowledges the connection between algorithmic reliability and data rights. However, the Act does not provide an explicit right to explanation or require algorithmic impact assessments for non-SDF entities.

C. The IP-Audit Collision

The conflict between audit mandates and trade secret protection is most pronounced when an independent data auditor, appointed by an SDF under Section 10(2)(b), seeks access to model internals to assess whether the LLM's processing of personal data complies with the Act. A comprehensive *"white-box"* audit, which involves direct access to model weights, training data, and fine-tuning pipelines, would require disclosure of information that LLM developers consider their most valuable trade secrets. The auditor's report, submitted to the DPBI under Rule 13, could expose this sensitive information to regulatory staff, contractors, and future data subjects exercising access rights, thereby increasing the risk of information leakage.

IV. COMPARATIVE PERSPECTIVES: THE EU AI ACT AND BEYOND

A. The EU AI Act's Tiered Approach

The EU AI Act, Regulation (EU) 2024/1689, which entered into force on 1 August, 2024, represents the most advanced current framework for balancing algorithmic oversight with intellectual property protection.[15] The Act employs a risk-based classification: AI systems presenting unacceptable risk are prohibited; high-risk systems are subject to rigorous pre-deployment conformity assessments; limited-risk systems must comply with transparency requirements; and minimal-risk systems remain largely unregulated. General-purpose AI *(hereinafter referred to as "GPAI"*) models, which include most advanced LLMs, are subject to obligations related to systemic risk assessment, adversarial testing, and incident reporting.

Article 78 of the EU AI Act establishes explicit confidentiality protections for information obtained by national competent authorities during enforcement activities. Regulatory sandboxes created under Article 57 are subject to confidentiality provisions that prohibit public disclosure of

[15] Commission Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), 2024 O.J. (L 1689) 1.

exit reports unless the provider consents.[16] This regulatory structure illustrates that robust algorithmic oversight and intellectual property protection can coexist, as the EU's framework intentionally incorporates confidentiality safeguards within the audit process.

B. Black-Box Auditing: A Technical Framework

Recent computer science literature indicates that effective auditing of LLMs does not require access to model weights. Black-box auditing, in which the auditor interacts with the model exclusively through its input-output interface, has emerged as a technically feasible and legally preferable alternative to white-box inspection.

Researchers from the Curiosity-Driven Auditing for Large Language Models *(also known as "CALM")* project have shown that reinforcement learning-based audit agents can detect harmful and biased input-output pairs in LLMs under black-box conditions, without access to model parameters.[17] Additionally, the Rank-Based Uniformity Test *(also known as "RUT")* introduced in recent literature allows for verification of an LLM's behavioral equivalence to a certified reference model using only output-level observations, without requiring access to logits.[18] Legally, black-box auditing avoids trade secret concerns entirely, as auditors do not access information constituting the developer's protectable intellectual property.

Explainable AI (XAI) techniques, including saliency mapping, rule extraction, and counterfactual reasoning, provide a complementary method. Tools such as IBM's AIX360, Alibi from SeldonIO, and Explabox allow auditors to evaluate fairness and robustness at the output level, producing interpretable insights into model behavior without revealing architectural details.[19] Although XAI has limitations, as explanations may not fully represent the model's computational processes, these techniques offer regulators actionable information regarding discriminatory outputs, privacy-sensitive responses, and compliance failures, without risking exposure of trade secrets associated with white-box access.

---

[16] EU AI Act, *supra* note 14, arts. 57, 78.
[17] Xiang Zheng et al., *CALM: Curiosity-Driven Auditing for Large Language Models* (Jan. 2025), https://arxiv.org/pdf/2501.02997 (last accessed on June 14, 2026)
[18] *Auditing Black-Box LLM APIs with a Rank-Based Uniformity Test*, arXiv:2506.06975 (June 2025), https://arxiv.org/abs/2506.06975 (last accessed on June 14, 2026)
[19] *From Black Box to Glass Box: A Practical Review of Explainable Artificial Intelligence (XAI)*, MDPI AI (Nov. 3, 2025), https://www.mdpi.com/2673-2688/6/11/285 (last accessed on June14, 2026)

V. A PROPOSED AUDIT FRAMEWORK FOR INDIAN DATA PROTECTION AUTHORITIES

A. Foundational Principles

The proposed framework is grounded in four foundational principles. The first is, **Proportionality** - regulatory access to proprietary model information should correspond to the sensitivity of the LLM use case and the potential magnitude of harm to data subjects, in line with the proportionality doctrine established in Puttaswamy.[20] The second is, **Confidentiality By Design** - information obtained by the DPBI or independent auditors during audits should be protected by statutory confidentiality obligations equivalent to those in Article 78 of the EU AI Act, with penal consequences for breaches. The third is, **Technical Pluralis** - the framework should support multiple audit methodologies, including black-box testing, explainable AI (XAI) techniques, documentation review, and process audits, without requiring white-box access as a default. The fourth is, **Iterative Adequacy** - audit standards should be periodically updated by the DPBI, in consultation with technical experts and industry, to reflect advances in both LLM capabilities and audit methodologies.

B. Tiered Access Model

The framework introduces a three-tier access model aligned with the risk profile of each LLM deployment:

Tier 1: Documentation and Process Audit - This tier applies to all SDF-deployed LLMs. Auditors review training data governance documentation (without direct inspection of the data), model cards, bias assessment reports, consent management records, and incident logs. No direct access to model internals is permitted, and there is no exposure of intellectual property.

Tier 2: Black-Box Behavioral Testing - This tier is triggered when a Tier 1 audit identifies potential indicators of bias, privacy violations, or non-compliance.[21] Trained audit agents, either

[20] *AI and Machine Learning Under DPDPA: Compliance Guide for Organizations*, DPDPA.com (Feb. 1, 2026), https://www.dpdpa.com/blogs/ai_machine_learning_dpdpa_compliance_guide.html (last accessed on June 14, 2026).
[21] *AI Auditing and the Access Question: Exploring Black-box Methods*, https://shcen.github.io/assets/files/ai_auditing.pdf (last accessed on June 14, 2026).

from the DPBI or independent auditors under contract, interact with the LLM via its Application Programming Interface *(also known as "API")*, conducting structured prompt campaigns to assess for discriminatory outputs, personal data memorization, and safety failures. All test inputs, outputs, and findings are protected by statutory confidentiality. Developers retain the right to contest findings and request re-testing prior to any adverse determination.

Tier 3: Supervised White-Box Review - This tier is reserved for high-risk deployments, such as LLMs used in criminal justice, credit scoring, or medical diagnosis, and is initiated only when Tier 2 testing reveals systemic compliance failures that cannot be adequately assessed without internal access. Access is granted within a regulatory sandbox environment, with the developer's legal counsel present, and under a statutory confidentiality agreement that binds the DPBI and all audit personnel. Model weights and training data are accessed solely to the extent necessary to characterize the identified failure; these materials may not be copied, retained, or shared outside the sandbox. Developers retain the right to seek a stay of Tier 3 access through application to the High Court where judicial review is pending.

C. Regulatory Sandbox Architecture

The Ministry of Electronics and Information Technology (MeitY)'s India AI Governance Guidelines, released in November 2025[22], explicitly endorse regulatory sandboxes as mechanisms for innovation and risk mitigation, identifying them as central to India's responsible AI framework. The proposed audit framework would establish a dedicated AI Audit Sandbox within the DPBI's technical infrastructure. This sandbox would: (i) provide air-gapped computing environments to prevent external access to model data during Tier 3 audits; (ii) generate cryptographically signed audit logs accessible to both the DPBI and the developer; (iii) employ third-party technical auditors who have passed a background-verification protocol; and (iv) produce findings in a standardized format that characterizes compliance failures without disclosing proprietary technical information.

D. Amendments to the DPDP Framework

The proposed framework necessitates targeted legislative and regulatory amendments. First, Section 10 of the DPDP Act should be amended to explicitly authorize the DPBI to conduct

[22] Ministry of Elecs. & Info. Tech., India AI Governance Guidelines (Nov. 5, 2025), https://www.pib.gov.in/PressReleasePage.aspx?PRID=2186639 (last accessed on June14, 2026).

black-box audits directly and to access regulatory sandboxes for Tier 3 reviews, clarifying that such access constitutes a legitimate exercise of the Board's inspection powers under Section 28(7).[23] Second, the DPDP Rules should be amended to include Rule 14, modelled on Article 78 of the EU AI Act, that imposes statutory confidentiality obligations on all DPBI personnel, independent auditors, and contractors who access proprietary model information during audits. Breach of this obligation should be classified as a cognizable offence. Third, MeitY should issue sector-specific guidance on audit standards for LLMs deployed in high-risk sectors, drawing on NITI Aayog's Principles for Responsible AI and aligning with emerging ISO 42001 (AI Management System) standards.[24]

E. The Role of Model Cards and Algorithmic Impact Statements

A practical supplement to the tiered audit model is the introduction of mandatory requirement for model cards and Algorithmic Impact Statements *(also known as "AISs")* for all SDF-deployed LLMs. Based on documentation practices endorsed by major AI developers and referenced in the EU AI Act's transparency obligations, model cards would require disclosure of intended use cases and known limitations, training data categories (without revealing precise composition), bias evaluation results on standard benchmarks, and aggregated red-teaming findings. The AIS, analogous to a DPIA but focused on societal rather than individual impacts, would assess the potential for discriminatory outcomes, privacy harms, and systemic risks. The AIS would be filed with the DPBI at the time of SDF designation and updated annually. Neither the model card nor the AIS would require disclosure of model weights, architectural specifics, or training data, ensuring compatibility with trade secret protection while providing regulators and the public with meaningful accountability information.[25]

VI. CONCLUSION

The perceived conflict between trade secret protection for proprietary LLMs and the accountability requirements of India's data protection framework should be seen as a design

[23] *Transforming Data Privacy: DPDP Act, 2023 and DPDP Rules, 2025*, Ernst & Young India (Jan. 21, 2026), https://www.ey.com/en_in/insights/cybersecurity/transforming-data-privacy-digital-personal-data-protection-rules-2025 (last accessed on June 14, 2026).
[24] NITI Aayog, Approach Document for India Part 1: Principles for Responsible AI (Feb. 2021), https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf (last accessed on June 14, 2026).
[25] *AI and Data Protection in India: DPDP Act 2023 Explained*, Complinity (Jan. 8, 2026), https://complinity.com/blog/compliance/ai-and-data-protection/ (last accessed on June 14, 2026).

challenge, not a binary opposition.[26] Well-structured audit frameworks can achieve the regulatory objectives of the DPDP Act, such as identifying bias, detecting privacy violations, and ensuring accountability, without requiring disclosure of model weights, architectures, or training data.

The three-tier audit model outlined in this article integrates documentation review, black-box behavioral testing, and supervised white-box sandbox access. It offers a proportionate, technically robust, and legally defensible framework for Indian data protection authorities. Drawing inspiration from the confidentiality provisions and regulatory sandbox mechanisms of the EU AI Act, this model is adapted to India's legal context, where trade secret protection is primarily contractual[27]. It also proposes specific legislative amendments to establish the framework within positive law.

As India advances toward comprehensive AI governance through initiatives such as the IndiaAI Mission and the MeitY AI Governance Guidelines, 2025, there is an opportunity to establish a global standard for algorithmic oversight that respects intellectual property[28]. The foundational constitutional commitment to privacy articulated in Puttaswamy necessitates such an approach.

[26] *Artificial Intelligence and Trade Secrets*, Cyber Blog India (Apr. 6, 2026), https://cyberblogindia.in/artificial-intelligence-and-trade-secrets/ (last accessed on June 14, 2026).
[27] *India's New Digital Privacy Law: The DPDP Act and Rules Explained*, Lexology (Dec. 1, 2025), https://www.lexology.com/library/detail.aspx?g=f4b2877e-a81b-4286-bdbf-77c4d9fd0807 (last accessed on June 14, 2026).
[28] *India AI Governance Guidelines: Empowering Ethical and Responsible AI*, IndiaAI (Nov. 5, 2025), https://indiaai.gov.in/article/india-ai-governance-guidelines-empowering-ethical-and-responsible-ai (last accessed on June 14, 2026).