ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

Explainable AI for Financial Fraud Detection in India: A Conceptual Framework Integrating Machine Intelligence, Chartered Accountancy Judgement, and the Emerging Indian Regulatory Architecture

Dr. Mahesha C[1], Dr. Naveen Kumar C.G[2]

Department of Commerce, Karnataka State Open University, Mysuru[1]

Department of Computer Science, Karnataka State Open University, Mysuru[2]

Abstract: Indian financial institutions now deploy machine-learning systems at unprecedented scale — the Unified Payments Interface alone processes transaction volumes that make manual fraud surveillance infeasible — yet the opacity of these models sits uneasily beside three distinctly Indian demands: the engineering imperative for interpretable outputs; the chartered accountancy profession's standards of professional scepticism codified in SA 240; and a regulatory architecture that has shifted from advisory soft law (NITI Aayog's Responsible AI principles) to binding instruments (the Digital Personal Data Protection Act 2023, the RBI Digital Lending Directions 2025, and SEBI's 2025 amendments assigning responsibility for AI-tool outputs). The scholarly literature addresses these demands separately: explainable artificial intelligence (XAI) research optimises explanation fidelity without engaging Indian audit practice; accountancy research on AI adoption in India remains descriptive; and compliance commentary analyses each instrument in isolation. This article develops TEA-India — an adaptation of the Tripartite Explainability Alignment framework — comprising a technical-explainability layer, a judgement-alignment layer anchored in ICAI's Standards on Auditing, and a compliance layer mapped to India's sector-specific instruments, connected through an explanation audit trail. The analysis identifies five India-specific gaps, including the absence of any automated-decision provision comparable to GDPR Article 22 in the DPDP Act, the consultative (not yet binding) status of RBI's model-risk guidance, and the near-complete absence of Indian empirical research on how chartered accountants actually consume model explanations. Five propositions and a comparative analysis against global frameworks are advanced. The framework offers Indian banks, NBFCs, auditors, and regulators a common vocabulary for designing fraud-detection systems whose explanations are simultaneously technically faithful, professionally usable, and legally defensible.

Keywords: Explainable artificial intelligence, financial fraud detection, Indi; chartered accountancy, professional skepticism, DPDP Act; RBI, SEBI, SA 240, human oversight

1. INTRODUCTION

India's financial system has digitised faster than any comparable economy. The Unified Payments Interface (UPI), Aadhaar-linked authentication, and account aggregator infrastructure have brought hundreds of millions of citizens into formal finance, and with them a fraud surface that no purely human control system can police. Card fraud, UPI-linked social-engineering scams, and digital-lending fraud are now detected overwhelmingly by machine-learning models operating inside banks and non-banking financial companies (NBFCs). These models score transactions in milliseconds, freeze suspicious credit lines, and flag accounts for investigation — decisions whose internal logic is opaque to every person downstream of them: the fraud analyst in a Mumbai operations centre, the statutory auditor signing off on internal financial controls under the Companies Act 2013, and the banking regulator demanding justification for customer-level action.

That opacity creates friction at three boundaries that are structurally the same as in other jurisdictions but institutionally distinct in India. The first boundary is technical. Post-hoc explanation methods such as LIME (Ribeiro et al., 2016) and SHAP (Lundberg & Lee, 2017) can approximate a model's local behaviour, but their fidelity and stability on the noisy, high-dimensional, code-mixed data of Indian retail finance remains largely unexamined (Burrell, 2016; Guidotti et al., 2018). The second boundary is professional. Indian statutory audit is governed by the Institute of Chartered Accountants of India (ICAI), whose Standards on Auditing are adapted from international standards; SA 240 (ICAI,

2009) obliges the auditor to exercise professional scepticism with respect to fraud, including presumed risks of revenue recognition and management override of controls. An auditor confronting a neural-network fraud model must translate feature-attribution scores into the language of misstatement risk, internal financial controls, and the fraud risk factors that SA 240 contemplates — and no standard currently instructs how. The third boundary is regulatory, and it is here that India diverges most sharply from the European Union template that dominates the XAI-governance literature. India has no omnibus AI statute. Instead, obligations arrive sectorally: the Digital Personal Data Protection Act 2023 (DPDP Act) creates consent, notice, and grievance-redressal duties but — unlike the GDPR — no explicit right against solely automated decisions; the Reserve Bank of India (RBI) polices lending conduct through the Digital Lending Directions 2025 and has consulted on model-risk management principles that would, if finalised, follow the logic of the US SR 11-7 guidance (Reserve Bank of India, 2024, 2026); and the Securities and Exchange Board of India (SEBI) has moved early, through its 2025 amendments, to assign explicit responsibility for the outputs of AI tools used by market infrastructure institutions and regulated intermediaries (SEBI, 2024, 2025).

The existing literatures fail to meet each other at these boundaries. XAI research is largely developed and benchmarked on Western datasets, with Indian applications emerging only recently (Hazarika et al., 2025) and rarely tested against Indian regulatory vocabulary. Accountancy research in India has documented the profession's AI readiness and ethics concerns descriptively, without formalising what explanation forms would actually support SA 240-compliant scepticism (Munoko et al., 2020, provides the ethics template; Indian adaptations remain scarce). Regulatory commentary, meanwhile, analyses the DPDP Act, RBI directions, and SEBI circulars as separate silos, even though a single fraud model at a large Indian bank is simultaneously subject to all three — plus the audit evidence standards of the Companies Act framework.

This article addresses that fragmentation for the Indian context. Its objective is to adapt and extend the Tripartite Explainability Alignment (TEA) framework into TEA-India: an integrated architecture in which technical explainability choices, chartered-accountancy judgement processes, and India's sector-specific regulatory obligations are treated as co-determining design constraints rather than independent afterthoughts. The contribution is threefold. First, the paper synthesises the Indian regulatory instruments and maps them onto a unified explanation-compliance architecture — something no published work has yet attempted. Second, it identifies five India-specific research gaps, including the DPDP Act's silence on automated decisions, the draft status of RBI model-risk guidance, and the absence of Indian behavioural evidence on auditor consumption of explanations. Third, it advances five propositions that convert those gaps into a testable research programme for Indian institutions and scholars.

The remainder proceeds as follows. Section 2 states the research questions. Section 3 reviews the three literatures with emphasis on their Indian development. Section 4 describes the conceptual methodology. Section 5 synthesises the India-specific gaps. Section 6 presents TEA-India. Section 7 develops the propositions and regulatory mapping. Section 8 compares TEA-India with the EU/US-anchored predecessor. Sections 9–11 discuss implications, limitations, future research, and conclusions.

2. RESEARCH QUESTIONS

RQ1. What forms of technical explanation do machine-learning fraud-detection systems deployed in Indian financial institutions currently afford, and how well do they serve the cognitive demands of professional scepticism under ICAI's SA 240?
RQ2. What explanation-related obligations arise for AI-driven fraud detection under the DPDP Act 2023, the RBI's digital lending and model-risk instruments, SEBI's AI-responsibility framework, and NITI Aayog's Responsible AI principles — and where do these instruments overlap, conflict, or remain silent?
RQ3. What conceptual architecture would allow Indian financial institutions to align XAI design, chartered-accountancy judgement, and the sectoral Indian regulatory framework within a single explainable fraud-detection system?

3. LITERATURE REVIEW

3.1 Machine-Learning Fraud Detection and the Indian Opacity Problem

The global trajectory of fraud-detection research — from logistic regression and expert systems through ensembles and deep learning — is well catalogued (Ngai et al., 2011; Bahrammirzaee, 2010), and Indian practice follows it. Large Indian banks and payment processors operate gradient-boosted ensembles over transaction streams; fintech lenders embed fraud scores in credit decisioning; and recent Indian work extends this frontier to privacy-preserving, explainable, real-time scoring (Hazarika et al., 2025; Sundaravadivel, 2026). The performance logic is compelling: at UPI-scale volumes, only statistical models can scan every transaction in real time.

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

The opacity problem, however, is at least as acute in India as elsewhere, and Burrell's (2016) three forms of opacity — proprietary secrecy, technical illiteracy, and intrinsic scale-driven complexity — all apply with local force. Fraud models are treated as trade secrets by competing lenders; the analysts and chartered accountants downstream of them are rarely trained to interrogate them; and the models themselves are no less inscrutable for being deployed in rupee denominations. Rudin's (2019) argument that post-hoc explanations of black boxes are unreliable precisely where stakes are highest therefore lands with full weight on Indian fraud systems that trigger card blocks, loan rejections, and account freezes. What distinguishes India is not the opacity but the accountability environment now forming around it — sectoral, enforcement-oriented, and increasingly explicit about human oversight (RBI, 2026).

3.2 Explainable Artificial Intelligence Methods

The XAI method stock is jurisdiction-neutral: locally faithful surrogates (Ribeiro et al., 2016), Shapley-value attribution (Lundberg & Lee, 2017), rule-based anchors, counterfactuals, and intrinsically interpretable model classes, organised in the taxonomies of Adadi and Berrada (2018) and Barredo Arrieta et al. (2020). Critiques are equally portable: instability of post-hoc explanations (Guidotti et al., 2018), the social and contrastive nature of human explanation (Miller, 2019), and the perils of explanation theatre (Burrell, 2016).

Two India-specific qualifications are nonetheless needed. First, benchmark evidence: the fraud datasets on which XAI methods are typically validated (e.g., European card-fraud corpora) differ materially from Indian transaction data in fraud-base rates, channel mix (UPI and wallets rather than cards), and linguistic noise in text fields; explanation methods tuned elsewhere cannot be assumed to transfer without local validation — an empirical question Indian scholarship has barely opened (Hazarika et al., 2025). Second, deployment economics: Shapley computation and explanation storage at UPI-scale volumes impose costs that shape which methods are feasible for Indian institutions, a constraint the global XAI literature treats lightly and Indian banks cannot.

3.3 Chartered Accountancy Judgement, Professional Scepticism, and AI in India

The international judgement literature provides the theoretical spine. Technology dominance theory holds that decision aids can progressively displace professional judgement (Arnold & Sutton, 1998); automation bias and complacency demonstrate the empirical reality of that displacement (Skitka et al., 1999; Parasuraman & Manzey, 2010); and algorithm aversion shows that observing a model err can cause professionals to abandon it even when it outperforms them (Dietvorst et al., 2015). Kokina and Davenport (2017) argue that automation redistributes rather than removes judgement; Preece (2018) insists that explanations must be usable by the professionals who act on them; Appelbaum et al. (2017, 2018) document the persistent deficit in converting machine outputs into audit evidence.

In India, this literature meets the institutional reality of the ICAI. SA 240, "The Auditor's Responsibilities Relating to Fraud in an Audit of Financial Statements" (ICAI, 2009), mirrors ISA 240: the auditor maintains professional scepticism throughout the audit, identifies and assesses risks of material misstatement due to fraud, and responds with procedures addressing presumed risks — improper revenue recognition and management override among them. Indian statutory audits additionally operate under the Companies Act 2013, whose section 143 obliges auditors to report fraud to the central government or audit committee in prescribed circumstances — a reporting duty that makes the provenance and interpretability of any AI-generated audit evidence materially consequential for the auditor's own legal exposure. The behavioural evidence on how Indian chartered accountants consume model explanations is, however, essentially absent: the Indian AI-in-audit literature is dominated by readiness surveys and syllabus debates rather than controlled studies of explanation use. This absence is itself a finding — the judgement layer of any Indian explainable-fraud framework currently rests on theory imported from Western experiments, not on evidence about Indian professionals.

3.4 The Indian Regulatory Architecture for AI in Financial Services

India's regulatory response to AI has been iterative and sectoral, and its trajectory matters for any explainability framework. The earliest instrument is soft law: NITI Aayog's *Responsible AI: Approach Document for India, Part 1 — Principles for Responsible AI* (NITI Aayog, 2021), which proposed seven principles — safety and reliability, equality, inclusivity and non-discrimination, privacy and security, transparency, accountability, and the protection and reinforcement of positive human values — explicitly grounded in the Indian constitutional context. These principles are advisory, but they established transparency as the reference point for subsequent hard law.

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

The DPDP Act 2023 (Government of India, 2023) is India's first binding cross-sector data statute. For explainable fraud detection, its provisions cut both ways. On the one hand, it obliges Data Fiduciaries to give notice of processing, honour consent, provide access, correction, and erasure rights, and operate grievance-redressal mechanisms; the DPDP Rules, 2025 (Government of India, 2025) were notified to operationalise the Act, with the Data Protection Board established and main provisions brought into force from 13 November 2025. On the other hand, and decisively for this paper, the DPDP Act contains no analogue to GDPR Article 22: there is no enumerated right of data principals not to be subject to solely automated decision-making, and therefore no statutory hook that directly forces explanation of an algorithmic fraud decision. Instead, the Act's Schedule carves out processing "for the purpose of ascertaining the financial information and assets and liabilities of any person who has defaulted in payment" — a fraud-adjacent exemption — and its penalty provisions (financial penalties for breach of fiduciary duties, with the Act setting penalty magnitudes that reach ₹500 crore for certain failures of security safeguards) create an accountability backstop rather than an individual explanation right.

Within its regulatory perimeter, the RBI has been more direct. The RBI (Digital Lending) Directions, 2025 (Reserve Bank of India, 2025) consolidate and supersede the 2022 digital-lending guidelines, requiring regulated entities to conduct due diligence on digital lending applications, ensure that credit decisions remain with the regulated entity rather than the lending-service provider, and maintain auditable records of the lending process. Model-level governance is addressed by consultative instruments: the draft "Regulatory Principles for Management of Model Risks in Credit" (RBI, 2024), followed by the broader draft "Guidance on Regulatory Principles for Model Risk Management" (RBI, 2026), which proposes board-approved model risk management frameworks, model inventories, independent validation, tiered risk oversight, human oversight of AI-driven decisions, and even a "kill switch" concept for AI models — a close cousin of the US SR 11-7 architecture, though still in draft and open for comment rather than binding.

SEBI has moved fastest to hard law. A consultation paper of 13 November 2024 (SEBI, 2024) proposed amendments assigning responsibility for the use and outputs of AI tools by market infrastructure institutions and regulated intermediaries; the amendments were subsequently notified, making regulated entities fully responsible for AI-tool outputs regardless of whether the tools are proprietary or third-party (SEBI, 2025), with a further 2025 consultation on guiding principles for responsible AI/ML usage in securities markets. The cumulative effect is an Indian architecture that is fragmented across instruments, uneven in force (statute; directions; draft guidance; notified amendments), but convergent in direction: transparency, human accountability, and auditability of algorithmic decisions.

4. METHODOLOGY

This article employs conceptual synthesis appropriate to a theoretical-contribution study, adapted from the methodology of the parent TEA framework. No claim is made of a systematic review: the corpus was assembled by snowball sampling from the verified XAI and accountancy literatures of the parent study (all DOIs re-verified against the Crossref REST API), supplemented by primary Indian regulatory instruments consulted at their official sources (the DPDP Act and Rules at the Gazette; RBI directions and draft guidance at rbi.org.in; SEBI circulars at sebi.gov.in; NITI Aayog documents at niti.gov.in; SA 240 at ICAI publications). Regulatory claims were verified against the texts of the instruments themselves rather than secondary commentary, and the status of each instrument (in force, notified, draft, consultation) was recorded at verification time, since the Indian framework is actively evolving and misdating an instrument's status is itself a factual error. The synthesis then triangulated the three literatures to derive the India-specific gaps, framework adaptation, and propositions. This method claims adequacy, not exhaustiveness, and its limitations are discussed in Section 9.

5. RESEARCH GAP (INDIA-SPECIFIC)

Five gaps emerge at the intersections of the Indian literatures and instruments.

Gap 1: Explanation–Risk-Factor Misalignment under SA 240. XAI methods emit feature attributions; SA 240 reasoning proceeds through fraud risk factors, presumed risks, and management override. No Indian standard, ICAI guidance note, or research protocol currently specifies how a SHAP vector on a transaction model is to be converted into SA 240-compliant audit evidence — an evidentiary gap the Companies Act fraud-reporting duty makes consequential (ICAI, 2009).

Gap 2: No Automated-Decision Explanation Right in the DPDP Act. Unlike the GDPR's Article 22, the DPDP Act grants no right against solely automated decisions and therefore no derived explanation entitlement at the individual level; grievance redressal and notice duties partially compensate, but an aggrieved fraud-flagged customer's entitlement to know "the logic involved" has no statutory anchor in India (Government of India, 2023). Explainability in India is thus a supervisory expectation rather than an individual right — a structural difference from the EU that frameworks imported wholesale will misrepresent.

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

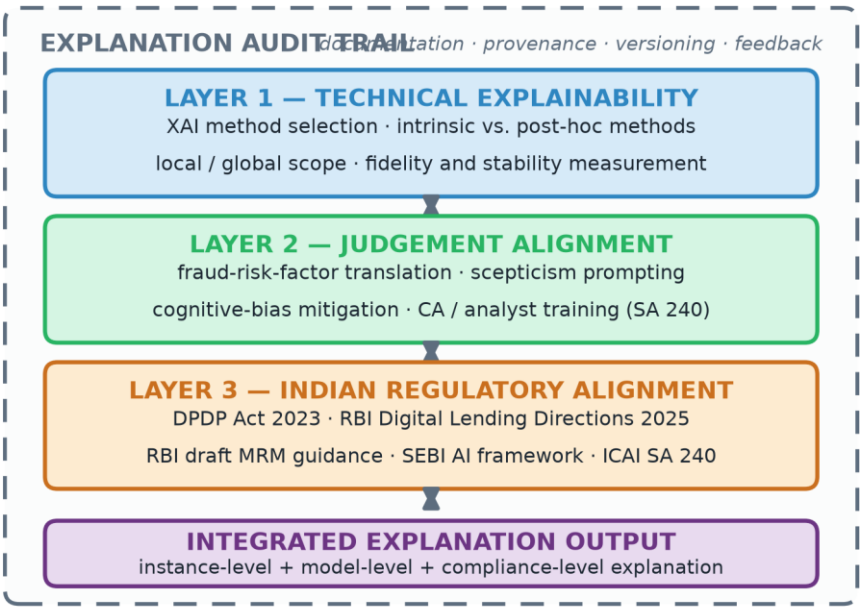
Gap 3: The Draft-Law Gap in Model Governance. RBI's model-risk architecture (validation, inventories, human oversight, kill switch) exists today only in draft guidance (RBI, 2024, 2026). Indian institutions designing fraud-model governance now face a moving target: SEBI's responsibility rules are binding for securities entities, while the banking model-risk rules they most resemble remain aspirational. No framework tells institutions how to design once, ahead of finalisation, for all instruments.

Gap 4: Absence of Indian Behavioural Evidence. The judgement layer of any explainable-fraud architecture rests on automation-bias and algorithm-aversion findings obtained almost entirely in Western laboratories (Parasuraman & Manzey, 2010; Dietvorst et al., 2015). There is no published experimental evidence on how Indian fraud analysts or chartered accountants actually weigh model explanations — a foundational evidentiary gap for Layer 2.

Gap 5: Sectoral Fragmentation with No Cross-Instrument Map. A fraud model at a diversified Indian financial group touches the DPDP Act, RBI directions and draft model-risk guidance, SEBI rules if the group has securities entities, NITI principles as soft-law reference, and ICAI standards for its statutory auditors. No published work maps these obligations onto a single explanation architecture; each commentary treats one instrument alone.

0.7. 6. Main Research Contribution: The TEA-India Framework

TEA-India adapts the parent TEA architecture (Figure 1) to the Indian context. It retains three interdependent layers joined by an explanation audit trail, but re-anchors the compliance layer in Indian instruments and the judgement layer in ICAI's standards.



The TEA-India framework for explainable AI in financial fraud detection. Layer 1 specifies the technical explanation affordances; Layer 2 translates them into SA 240-compliant professional scepticism; Layer 3 maps them onto the Indian regulatory architecture (DPDP Act 2023, RBI Digital Lending Directions 2025, RBI draft model-risk guidance, SEBI AI framework, ICAI SA 240). The explanation audit trail links instance-level explanations to model-level documentation and compliance evidence.

Figure 1. The TEA-India framework. The three layers are interdependent: technical choices constrain judgement alignment, judgement requirements shape acceptable methods, and both are bounded by India's sectoral regulatory obligations. The explanation audit trail maintains versioned records linking each explanation to model documentation and to compliance evidence across the DPDP Act, RBI, SEBI, and ICAI frameworks.

Layer 1 — Technical Explainability. The layer retains the parent's method stock — intrinsic models preferred for high-stakes decisions per Rudin (2019), post-hoc local attributions where performance demands — but adds two India-specific design constraints: local validation of explanation fidelity on Indian transaction data (channel mix, fraud-base rates, code-mixed text), and deployment-cost realism at UPI-scale volumes. Explanation fidelity remains the primary quality metric (Guidotti et al., 2018).

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

Layer 2 — Judgement Alignment (Chartered Accountancy). Explanations must be rendered in the conceptual vocabulary of SA 240: a feature attribution on transaction velocity becomes an assertion about opportunity and pressure in the fraud-triangle framing; a model's risk score on an entity becomes assessed risk of material misstatement with documented rationale. The layer incorporates bias-mitigation interface design (presenting base rates alongside model outputs to counter automation complacency; Parasuraman & Manzey, 2010) and extends to the Companies Act context: because section 143 makes fraud reporting consequential for the auditor, the audit trail must record how the professional evaluated, accepted, or challenged each machine flag — evidence of scepticism exercised, not merely asserted.

Layer 3 — Indian Regulatory Alignment. Each explanation output is mapped to the obligations it discharges across instruments: notice and grievance-redressal coherence under the DPDP Act (Government of India, 2023); process auditability and regulated-entity accountability under the RBI Digital Lending Directions (RBI, 2025); forward compatibility with the draft model-risk architecture — independent validation, model inventory entries, human-oversight artefacts, kill-switch readiness (RBI, 2024, 2026); output responsibility for SEBI-regulated entities (SEBI, 2025); and alignment with the transparency and accountability principles of NITI Aayog (2021) as the soft-law reference point.

The Explanation Audit Trail is the connective tissue: a versioned record of input data, model version, XAI method, explanation output, judgement-alignment translation, reviewer action, and compliance categorisation. In the Indian setting it serves three simultaneous masters — RBI supervisory inspection, DPDP accountability, and the statutory auditor's evidence file — and is the mechanism through which one artefact can discharge obligations across the fragmented architecture.

6.2 Propositions

P1. Fraud-risk-factor-calibrated explanations (fraud-triangle framings aligned to SA 240 vocabulary) will improve the diagnostic value of fraud alerts for Indian chartered accountants and fraud analysts relative to uncalibrated feature-attribution outputs.
P2. Explanation interfaces that present model reasoning alongside empirical base rates will reduce automation bias among Indian fraud analysts relative to interfaces presenting model output alone — the Indian behavioural test of Parasuraman and Manzey's (2010) attentional model, currently unrun.
P3. In the absence of a DPDP analogue to GDPR Article 22, explanation adequacy in Indian fraud detection will be driven primarily by supervisory expectations (RBI, SEBI) rather than individual rights; consequently, frameworks that embed explanation standards into model governance ex ante (draft MRM principles) will outperform post-hoc grievance-redressal compliance in producing usable explanations.
P4. Forward-compatible governance — building explanation, validation, inventory, and kill-switch artefacts ahead of RBI's draft MRM finalisation — will lower total compliance cost for Indian institutions than instrument-by-instrument retrofitting, because the audit trail discharges DPDP, RBI, SEBI, and ICAI evidence duties simultaneously (the Indian test of the parent framework's P5).
P5. SEBI's assignment of output responsibility to regulated entities will function as the de facto explanation-forcing mechanism in Indian securities markets, generating stronger explanation practices there than in banking while RBI's model-risk guidance remains in draft — producing a natural cross-sector experiment on regulation-driven explainability.

7. ANALYSIS: MAPPING XAI METHODS AND INDIAN REGULATORY INSTRUMENTS

Table 1 maps XAI method families to Indian fraud-detection use cases, the judgement roles they support, and the regulatory anchors they serve.
Table 1. XAI Methods Mapped to Indian Fraud-Detection Judgement Roles and Regulatory Anchors

XAI Method Family	Indian Use Case	Judgement Role (SA 240 / Companies Act)	Indian Regulatory Anchor	Limitation
Intrinsic interpretable models (GAMs, shallow ensembles)	UPI transaction scoring; digital-lending fraud screens	Decision logic traceable to variables; supports risk-assessment documentation	RBI draft MRM validation (RBI, 2026); NITI transparency	Performance trade-off at extreme scale

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

XAI Method Family	Indian Use Case	Judgement Role (SA 240 / Companies Act)	Indian Regulatory Anchor	Limitation
LIME (local surrogates)	Case-level alert triage	Scepticism at the alert level; hypothesis generation	DPDP grievance-redressal coherence; SEBI output responsibility	Instability; unvalidated on Indian data
SHAP (Shapley attribution)	Ranking drivers of entity risk scores	Prioritising investigative procedures; documented rationale for audit files	RBI Digital Lending process auditability (RBI, 2025)	Computational cost at UPI volumes
Anchors / rule lists	Threshold conditions for flagging	Maps to SA 240 presumed-risk procedures	ICAI evidence standards	Rules may over-fit local segments
Counterfactuals	What-if assessment of declines	Adverse-outcome reasoning for lending decisions	RBI fair-treatment conduct expectations	Plausibility constraints
Model cards / FactSheets	Model governance documentation	Validation evidence for statutory auditors	RBI draft MRM inventory; SEBI accountability; NITI principles	Static; needs audit-trail linkage

Table 2 maps the Indian instruments to their explanation-relevant obligations and the TEA-India components that address them.

Table 2. Indian Regulatory Instruments Mapped to Explainability Obligations and TEA-India Components

Instrument	Status	Key Explainability-Relevant Provision	TEA-India Component
DPDP Act 2023; DPDP Rules 2025	In force (phased from 13 Nov 2025)	Notice, consent, access/correction/erasure, grievance redressal; penalty regime (up to ₹500 crore for certain failures); no automated-decision right	Layer 3 + audit trail (accountability backstop)
RBI (Digital Lending) Directions, 2025	In force	Due diligence on lending apps; credit decisions with the RE; auditable process records	Layer 3 (process auditability)
RBI draft Regulatory Principles for Model Risk Management	Draft (2026; comments closed)	Board-approved MRM framework; model inventory; independent validation; human oversight; AI kill switch	Layers 1 + 3 + audit trail (forward-compatible design)
SEBI AI-responsibility amendments (2025)	Notified	Regulated entities fully responsible for AI-tool outputs; further consultation on responsible AI/ML principles	Layer 3 + audit trail (output accountability)
NITI Aayog Responsible AI (2021)	Advisory	Seven principles including transparency and accountability	Reference standard for all layers
ICAI SA 240 (2009); Companies Act 2013 s.143	In force	Professional scepticism; fraud risk assessment; mandatory fraud reporting	Layer 2 (judgement alignment)

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

8. COMPARATIVE ANALYSIS: TEA-INDIA VERSUS THE EU/US-ANCHORED TEA

Table 3 compares the Indian adaptation with the parent framework, which was anchored in the GDPR, the EU AI Act, and US SR 11-7.

Table 3. Comparative Analysis of TEA-India and the Parent TEA Framework

Dimension	TEA (EU/US anchor)	TEA-India
Data-protection anchor	GDPR Art. 22: individual right against solely automated decisions; explanation logic	DPDP Act: no automated-decision right; grievance redressal and penalty backstop
AI-specific statute	EU AI Act risk tiers; fraud-detection carve-out from high-risk credit scoring	No omnibus AI statute; sectoral instruments (RBI, SEBI) plus soft law (NITI)
Model-risk governance	SR 11-7 (binding US supervisory guidance)	RBI draft MRM principles (consultative, not yet binding)
Securities dimension	Absent from parent	SEBI output-responsibility amendments (binding)
Audit anchor	ISA 240 (IAASB)	SA 240 (ICAI) + Companies Act s.143 fraud-reporting duty
Human oversight source	AI Act Art. 14 (for high-risk systems)	RBI draft MRM; SEBI accountability; NITI Principle
Dominant compliance logic	Rights-driven (individual explanation entitlements)	Supervision-driven (institutional accountability and auditability)
Key uncertainty	GDPR/AI Act interpretation	Finalisation and scope of RBI MRM guidance; DPDP enforcement practice

The comparison reveals a structural inversion: the parent framework presumes an individual-rights baseline that pushes explanations outward to affected customers, whereas the Indian architecture pushes explanations upward to supervisors, boards, and auditors. For framework design this is not a defect but a specification difference — the explanation audit trail, which in TEA serves accountability and validation, becomes in TEA-India the central compliance artefact, because it is the single object that all Indian instruments — DPDP accountability, RBI auditability, SEBI responsibility, ICAI evidence — can be shown to converge upon.

9. DISCUSSION, IMPLICATIONS, AND LIMITATIONS

9.1 Theoretical and Practical Implications

Theoretically, TEA-India demonstrates that explainability frameworks are not portable as-is: the layer structure survives translation, but the compliance logic inverts from rights-driven to supervision-driven, and the judgement layer must be re-anchored in ICAI standards and the Companies Act's reporting duties. The framework also converts India's regulatory fragmentation into an analytic asset: because SEBI's rules are binding while RBI's model-risk guidance remains draft, P5 predicts — and the next few years will observe — a natural experiment in regulation-driven explainability across sectors.

Practically, the framework offers three audiences a common vocabulary. For Indian banks and NBFCs, it is a design checklist: every explanation capability assessed for technical fidelity, SA 240 usability, and forward compatibility with draft MRM obligations. For ICAI and statutory auditors, it specifies what to demand of client management's fraud models — not the model's proprietary internals, but the audit trail linking each alert to its explanation, its reviewer's action, and its compliance categorisation. For RBI and SEBI, it illustrates how a single mandated artefact — the explanation audit trail — could reduce duplicative compliance across instruments.

9.2 Limitations

The article shares the parent framework's limitations — conceptual, not empirical; propositions untested — and adds Indian ones. The regulatory corpus is a snapshot: the RBI model-risk guidance was in draft and the DPDP Rules recently notified at the time of writing, and any finalisation may alter the compliance mapping in Table 2. Behavioural claims about Indian professionals rest on imported theory (Gap 4); the framework predicts they will transfer, but prediction is not evidence. The UPI-scale cost analysis is qualitative; no Indian institution's implementation data were available. And the snowballed corpus, though verified at every citation, cannot claim systematic coverage of Indian-language scholarship or grey regulatory commentary.

ISSN Online 3108-026X, ISSN Print 3139-4833
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 11, November 2026 | DOI 10.17148/IMRJR.2026.031103

10. FUTURE RESEARCH

The five gaps specify the research programme. Gap 1 calls for ICAI-anchored protocol development: converting XAI outputs into SA 240 evidence templates, testable through case studies with practising chartered accountants. Gap 2 invites legal-scholarly analysis of whether DPDP grievance-redressal and sectoral conduct rules can functionally substitute for an automated-decision explanation right — and what their failure to do so implies for DPDP reform debates. Gap 3 requires implementation research with Indian institutions building forward-compatible governance ahead of RBI finalisation. Gap 4 demands the missing behavioural experiments: Indian fraud analysts and auditors as subjects, local fraud scenarios as stimuli, explanation interfaces as treatments — the Indian replication of Dietvorst et al. (2015) and Parasuraman and Manzey (2010) that the framework's P1–P2 require. Gap 5 calls for comparative sectoral studies as SEBI's binding rules and RBI's draft guidance diverge in force — the natural experiment of P5, and a rare opportunity to observe regulation-driven explainability with identifiable treatment and control groups.

11. CONCLUSION

India's fraud-detection systems operate at a scale, within a professional structure, and under a regulatory architecture that no imported framework fully describes. This article has adapted the Tripartite Explainability Alignment framework for that setting: technical explainability choices, chartered-accountancy judgement under SA 240 and the Companies Act, and a sectoral regulatory regime spanning the DPDP Act, RBI digital-lending and model-risk instruments, SEBI's AI-responsibility rules, and NITI Aayog's principles — joined by an explanation audit trail that serves them all. The analysis shows that explainability in India is structurally supervision-driven rather than rights-driven, that the sectoral fragmentation of Indian regulation converges on auditability as the common obligation, and that the nation's draft-versus-notified asymmetry creates both a compliance design problem and a research opportunity. Five propositions convert these observations into a testable programme. Whether Indian professionals perform better with calibrated explanations, whether forward-compatible governance lowers compliance cost, and whether binding rules generate better explanations than draft ones — these are now empirical questions with a framework waiting to answer them.

REFERENCES

[1]. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
[2]. Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big Data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory*, *36*(4), 1–27. https://doi.org/10.2308/ajpt-51684
[3]. Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2018). Analytical procedures in external auditing: A comprehensive literature survey and framework for external audit analytics. Journal of Accounting Literature, 40(1), 83–101. https://doi.org/10.1016/j.acclit.2018.01.001
[4]. Arnold, V., & Sutton, S. G. (1998). The theory of technology dominance: Understanding the impact of intelligent decision aids on decision maker's judgments. *Advances in Accounting Behavioral Research*, *1*(3), 175–194.
[5]. Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, *19*(8), 1165–1195. https://doi.org/10.1007/s00521-010-0362-z
[6]. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
[7]. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512
[8]. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033
[9]. Government of India. (2023). *The Digital Personal Data Protection Act, 2023* (Act No. 22 of 2023). The Gazette of India.
[10]. Government of India. (2025). The Digital Personal Data Protection Rules, 2025. Ministry of Electronics and Information Technology, The Gazette of India.
[11]. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009
[12]. Hazarika, I., Raghuveer, K., Jayanth, H., Tamilarasu, J., Kathiravan, C., & Rambabu, G. (2025). Adaptive and explainable machine learning framework for real-time credit scoring and financial fraud detection with privacy-preserving intelligence. *Proceedings of the 1st International Conference on Research and Development in*

Information, Communication, and Computing Technologies, 131–136. https://doi.org/10.5220/0013858800004919

[13]. Institute of Chartered Accountants of India. (2009). *Standard on Auditing (SA) 240: The auditor's responsibilities relating to fraud in an audit of financial statements*. ICAI.

[14]. Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting*, *14*(1), 115–122. https://doi.org/10.2308/jeta-51730

[15]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

[16]. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[17]. Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics*, *167*(2), 209–234. https://doi.org/10.1007/s10551-019-04407-1

[18]. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

[19]. NITI Aayog. (2021). *Responsible AI: Approach document for India, Part 1 — Principles for responsible AI*. Government of India.

[20]. Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

[21]. Preece, A. (2018). Asking 'Why' in AI: Explainability of intelligent systems — perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*, *25*(2), 63–72. https://doi.org/10.1002/isaf.1422

[22]. Reserve Bank of India. (2024). *Regulatory principles for management of model risks in credit* (Draft for comments, August 5, 2024). RBI.

[23]. Reserve Bank of India. (2025). *Reserve Bank of India (Digital Lending) Directions, 2025* (Master Direction, May 8, 2025). RBI.

[24]. Reserve Bank of India. (2026). *Draft guidance on regulatory principles for model risk management* (Press Release PR 2026-2027/528, June 24, 2026). RBI.

[25]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

[26]. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[27]. Securities and Exchange Board of India. (2024). *Proposed amendments with respect to assigning responsibility for the use of artificial intelligence tools by market infrastructure institutions, registered intermediaries and other persons regulated by SEBI* (Consultation Paper, November 13, 2024). SEBI.

[28]. Securities and Exchange Board of India. (2025). *Regulated entities responsible for output of AI usage* (Notified amendments and subsequent consultation on responsible AI/ML usage in securities markets). SEBI.

[29]. Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? International Journal of Human-Computer Studies, 51(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

[30]. Sundaravadivel, P. (2026). Federated deep learning and explainable AI for real-time credit card fraud detection in highly imbalanced transaction streams. *Discover Artificial Intelligence*, *6*(1). https://doi.org/10.1007/s44163-026-01449-z