Vol. 19 (2026): Proceedings of 2026 International Conference on Financial Technology, Mathematical Statistics, and Business Analytics (ICFMB 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/icfmb.202619402409

# Research on Algorithmic Explainability and Legal Liability Framework under AI Audit Background

**Wendong Li***

*Business School, Guilin University of Electronic Technology, Guilin, 541000, China*

**Corresponding author: Wendong Li.*

## Abstract

The deep integration of artificial intelligence (AI) in auditing has significantly enhanced audit efficiency and data processing capabilities. However, the inherent "black box" nature of algorithms poses severe challenges in determining liability for audit failures, which hinders further development of AI auditing technologies. This study aims to address liability attribution dilemmas caused by algorithmic opacity in AI auditing contexts. By integrating algorithm governance theory with extended audit accountability theory, we establish a bidirectional framework of "algorithmic explainability and legal liability." The research first analyzes the real-world contradiction between explainability deficiencies and legal lag in current AI auditing practices, revealing Deloitte's liability determination challenges stemming from generative AI hallucinations. Solutions are then proposed from both technical and legal dimensions. Findings indicate that algorithmic explainability serves as a prerequisite for legal liability determination, while clear legal regulations provide institutional safeguards for technological transparency. This framework not only clarifies accountability chains in human-machine collaboration but also offers theoretical support and practical guidance for promoting compliance in AI auditing and maintaining capital market trust mechanisms.

## Keywords

AI auditing, algorithmic explainability, legal liability, framework research, explainable artificial intelligence

## 1. Introduction

As a product of integrating artificial intelligence technology with traditional auditing practices, AI auditing is progressively reshaping the efficiency and accuracy of audit processes. However, issues stemming from the black-box nature of algorithmic decision-making and AI-induced illusions remain prevalent. With the application proportion of AI auditing tools in risk assessment and material misstatement detection core processes rising from 15% in 2020 to 42% in 2024, the interpretability of algorithm outputs has become a critical link connecting technological applications with audit accountability [1]. In audit practice, due to the complex logic of algorithmic models (such as deep learning neural networks) being difficult for auditors to comprehend, unclear audit definitions caused by algorithmic biases may lead to audit failures, while ambiguous liability attribution could trigger legal disputes and report credibility crises. On October 5, 2025, Deloitte's independent evaluation report for the Australian government contained numerous errors due to

generative AI usage, even citing non-existent literature. Rach, the first to identify the issues, stated: “When reports are based on flawed, initially undisclosed, and non-professional methodologies, you can no longer trust these recommendations.” Following such incidents, it becomes extremely challenging to determine liability attribution. Under traditional auditing frameworks, auditors primarily bear responsibility for failing to adequately identify fraud risks. However, this case involved the use of AI auditing tools (GPT-4o) with algorithmic outputs lacking interpretability. The core issue lies in whether the algorithm itself was flawed or Deloitte failed to conduct rigorous manual reviews. The auditing entity struggled to demonstrate its fulfillment of due diligence obligations during the audit process, leaving the question of accountability unresolved regarding the $291,000 refund to the Australian government. This case highlights that in AI auditing contexts, algorithmic interpretability—while a fundamental technical requirement—serves as a prerequisite for legal liability determination. Only when algorithmic decision-making processes are traceable, comprehensible, and verifiable can the boundaries of auditing entities' responsibilities in AI tool applications be clearly defined [2].

The widespread adoption of AI auditing represents an inevitable trend in the digital transformation of the auditing industry. AI-driven audit reforms require establishing a collaborative framework that balances algorithmic explainability with legal accountability to harmonize technological innovation and risk management. This study aims to enhance algorithmic explainability through technical improvements, clarify liability boundaries for stakeholders through legal frameworks, and define responsibility frameworks to delineate duties among algorithm developers, auditing institutions, and audited entities. By doing so, we seek to maximize the technical advantages of artificial intelligence while safeguarding the impartiality of the auditing profession and upholding the rigor of legal standards.

## 2. Core Concepts and Theoretical Foundations

### 2.1 Definition of Core Concepts

#### 2.1.1 AI Audit

AI auditing integrates artificial intelligence technologies with traditional auditing theories and methodologies. By employing algorithmic models, it automatically collects, identifies, cleans, analyzes, and processes financial data, operational metrics, and internal control master data of audit subjects, enabling data-driven analysis, risk assessment, and conclusion generation—a groundbreaking auditing paradigm. Characterized by intelligence, automation, standardization, and precision, it encompasses diverse audit services including financial statement auditing and internal control auditing. Compared to conventional manual auditing, AI auditing effortlessly overcomes human resource limitations, processes massive datasets, and enhances audit efficiency and accuracy [3]. However, it also faces challenges such as audit black boxes, information security risks for audited entities, and ambiguous accountability delineation.

#### 2.1.2 Algorithmic Explainability (XAI)

Algorithmic interpretability refers to the ability to explain an algorithm's decision-making process, rationale, and logic using human-understandable language [4]. Its core features involve breaking down algorithmic black boxes to achieve transparent and traceable decision-making. In AI auditing scenarios, this interpretability manifests through three key aspects: 1) Clear visualization and traceability of algorithms. Specifically, auditing algorithms should demonstrate data processing workflows, risk identification criteria, and logical frameworks for audit conclusions. Auditors can comprehend algorithmic decision-making mechanisms, while third-party oversight bodies can evaluate algorithmic rationality and compliance. Stakeholders can also assess the credibility of audit outcomes.

#### 2.1.3 Legal Liability for AI Auditing

AI audit legal liability refers to civil, administrative, or even criminal liabilities that may be incurred by responsible parties when using AI in auditing activities, resulting in erroneous audit conclusions or losses to stakeholders due to violations of laws, auditing standards, or contractual agreements. However, the complexity arises from multiple liable entities, diverse causative factors, ambiguous liability boundaries, and algorithmic decision-making uncertainties, making it challenging to delineate direct accountability boundaries and determine liability [5].

## 2.2 Theoretical Basis

### 2.2.1 Algorithm Governance Theory

Algorithm governance refers to a systematic management framework that employs technical, legal, and ethical approaches to regulate algorithm design, development, deployment, and usage, with the core objective of ensuring fairness, transparency, and accountability in algorithmic operations. This theory emphasizes that algorithms are not merely tools but decision-making mechanisms whose execution must comply with legal norms and social ethics [6]. In AI auditing scenarios, algorithm governance theory provides methodological guidance for enhancing algorithmic explainability and defining legal boundaries. It requires technical optimizations to break down algorithmic black boxes, clarify the responsibility boundaries of relevant stakeholders, and establish clear accountability frameworks.

### 2.2.2 Audit Responsibility Theory

The audit responsibility theory refers to the obligations that audit entities must fulfill during auditing activities, encompassing professional accountability and legal liabilities. Professional accountability requires auditors to adhere to auditing standards, maintain independence, obtain sufficient evidence, and deliver appropriate audit reports. Legal liabilities involve the consequences arising from negligence, fraud, or violations of laws and auditing standards that lead to audit failures. In AI-driven auditing, this theory should be expanded to incorporate algorithm developers and technology providers into the responsibility framework [7], clearly defining each party's obligations across different stages to achieve comprehensive audit accountability coverage.

## 2.3 Review on Algorithmic Explainability (XAI) and Legal Liability in AI Auditing

In summary, while existing research has achieved certain progress in algorithmic explainability (XAI) and audit legal liability, significant disciplinary fragmentation and practical disconnection remain unresolved. Current XAI studies predominantly focus on technical “black box” analysis, lacking alignment with professional audit judgment logic and legal evidentiary requirements. Meanwhile, audit liability research continues to rely on traditional human-to-human attribution models, struggling to address complex liability distribution challenges arising from algorithmic decision-making. This theoretical and technological lag makes it difficult to establish causal relationships between “algorithmic flaws” and “audit failures” when confronting emerging risks like generative AI hallucinations. Therefore, there is an urgent need to break down disciplinary barriers and establish a bidirectional interaction framework integrating technology and law to resolve the practical contradictions of insufficient algorithmic explainability and ambiguous legal liability definitions.

# 3. Current Status and Challenges of Algorithmic Explainability and Legal Liability under AI Audit

Deloitte's 2025 report “AI Reshaping Auditing: Exploring Industry Transformation Driven by Artificial Intelligence” indicates that generative AI applications could revolutionize the auditing sector, facilitating a transition from “copilot” to “agent” operational models. The most significant impact lies in reducing auditors' involvement in repetitive tasks, allowing them to focus more on professional judgment. [8] However, current challenges in the auditing field stem from both technological immaturity and systemic deficiencies in generative AI implementation.

## 3.1 Current Status and Challenges of Explainability in AI Review Algorithms

In AI auditing scenarios, algorithmic explainability requires auditors to clearly articulate input variables, decision-making logic, weight distribution, and output generation pathways of algorithm models, ensuring compliance with auditing standards and regulatory requirements. [2] However, in practical applications, this requirement is often difficult to achieve due to algorithms' inherent black-box characteristics, emergent properties, and cognitive differences between humans and machines. Consequently, data security concerns, user trust in machine models, and questionable rationality of algorithmic decisions hinder widespread adoption of AI auditing, making its outcomes hard to gain credibility among information users [10].

Interpretability and audit efficiency exhibit a trade-off relationship. Overemphasizing algorithmic interpretability may increase design and development costs while reducing audit efficiency, whereas excessive focus on audit efficiency could lead to insufficient algorithmic transparency. If high-precision yet low-interpretability algorithm models are selected during audits to prioritize efficiency, this may result in untraceable audit outcomes and make it difficult to identify responsible parties in case of audit failures.

## 3.2 Current Status and Challenges of Legal Liability in AI Auditing

The lack of algorithmic explainability in the context of AI auditing directly leads to legal liability determination dilemmas. With the widespread adoption of AI auditing, related legal disputes have been increasing. However, the lagging nature of existing legal norms and auditing standards poses significant challenges in defining AI auditing liabilities [11].

Throughout the AI audit process involving algorithm developers, auditing institutions, auditors, and audited entities, the identification of liability subjects remains ambiguous. Whether the failure stems from developers 'failure to define responsibilities during programming, auditors' non-compliance with audit procedures causing errors, or inaccuracies in information provided by audited entities—these liability determinations lack clear legal standards.

Due to the black-box nature of algorithmic auditing, establishing causal relationships between “algorithmic flaws-induced losses” and “auditing actions resulting in losses” becomes challenging when audit failures cause stakeholder damages. When losses occur due to auditing practices, it is difficult to determine the proportion of algorithmic failures versus auditors' negligence in the damage assessment, creating significant difficulties in establishing causal connections.

In current auditing practices, artificial intelligence (AI) has begun to be incorporated into audit processes with expanding applications. However, legal frameworks often lag behind technological advancements. Existing audit laws and standards primarily target traditional auditing methods, failing to account for the unique characteristics of AI auditing. They lack clear provisions regarding algorithmic explainability and liability allocation mechanisms, resulting in poor compatibility. For instance, the EU's Artificial Intelligence Act, which came into effect in August 2024, only addresses general AI applications while lacking dedicated legislation tailored to this specialized field [12].

In losses resulting from AI audit failures, information users are often the most direct victims. When infringement incidents occur, these users find themselves at a disadvantage, struggling to access core technical information of AI algorithms and records of decision-making processes, while algorithm developers and auditing institutions hold crucial evidence. The current allocation of burden of proof fails to account for this reality, leaving information users' legitimate rights and interests vulnerable when their interests are compromised.

## 3.3 Intrinsic Relationship Between Algorithmic Explainability and Legal Liability

The legal challenges surrounding the interpretability of AI audit algorithms fundamentally reflect the inherent tension between technological innovation and legal lag, which are closely interconnected. On one hand, enhanced algorithmic transparency provides crucial foundations for legal liability determination—only by ensuring decision-making processes are transparent and traceable can we clearly distinguish between responsible parties' actions and faults, thereby defining accountability boundaries. When AI algorithms demonstrate sufficient interpretability and clearly articulate their decision-making principles and logic, auditors can swiftly identify the root causes of errors. This enables precise differentiation between algorithmic design flaws, auditor negligence, or data falsification by audited entities, ultimately clarifying accountability. On the other hand, legal liability frameworks inherently drive algorithmic transparency improvements [13]. By legally specifying interpretability obligations for algorithm developers and auditing institutions, along with liability assessments for travel-related incidents, stakeholders are incentivized to optimize algorithmic transparency. This encourages greater technological investment in algorithmic black-box analysis and enhances audit algorithm development alongside disclosure practices.

# 4. Algorithmic Explainability and Legal Liability Framework Construction under AI Audit Background

## 4.1 Principles of Framework Construction

The audit responsibility framework developed in this study adheres to three core principles: compliance, practicality, and interconnectivity. It establishes that algorithmic audit outputs must strictly comply with auditing standards and the legal framework governing AI regulation as fundamental requirements. This framework ensures auditors' accessibility, regulatory traceability, and clear demarcation between interpretability and legal liability boundaries. Responsibility delineation is grounded in interpretability, while responsibility definition serves as a safeguard to enhance interpretability through accountability mechanisms.

## 4.2 Algorithm Interpretability Dimension Framework

### 4.2.1 Core Dimensions

Algorithm construction should prioritize explainability as the primary principle [14], requiring all AI audit algorithms to fully meet explainability requirements. Firstly, algorithmic explainability must permeate the entire audit process, encompassing all stages from algorithm design, development, deployment, application, to validation. Secondly, it should align with professional audit needs by providing clear, logically coherent support for audit conclusions, while logging algorithm inputs, feature selection, model computations, and output results to facilitate secondary verification by auditors and regulatory authorities. Thirdly, algorithmic explainability must strike a balance with audit efficiency—enhancing explainability without compromising operational efficiency [15]—to avoid excessive costs incurred by overemphasizing explainability.

### 4.2.2 Technical Approach

The preventive measure primarily addresses interpretability deficiencies and legal risks in AI audit algorithms. This can be achieved by establishing an algorithmic interpretability evaluation framework, with regulatory bodies and industry associations spearheading the development of corresponding testing systems and scoring mechanisms. Audit AI systems failing to meet predetermined thresholds should be denied approval, thereby creating industry-specific entry barriers. For self-developed AI audit systems, inherently interpretable models such as logistic regression and decision trees should be selected, while incorporating attention mechanisms and interpretability modules within deep learning architectures to ensure design-level compliance. Third-party AI audit tools must adopt interpretable solutions, with algorithm developers required to provide transparent API interfaces for interpretability verification.

In-process monitoring is primarily employed during AI audit implementation to ensure proper operation of explainability mechanisms and mitigate legal risks. By establishing a daily algorithm monitoring system, the operational processes of AI-based audit solutions are tracked, algorithm decision logs are recorded, and any unexplained issues or operational defects are promptly identified. These findings are immediately communicated to algorithm developers for optimization and corrective actions. The audit results generated by AI systems can be cross-validated through methods analogous to traditional sampling verification processes, ensuring the accuracy and credibility of audit conclusions.

LIME was employed for local explanation and SHAP for global explanation, with visualization tools integrated to present decision-making logic. Post-audit decisions should be generated promptly during the audit process to facilitate timely validation of decision rationality by the audit office.

### 4.2.3 Evaluation System

An evaluation index system was constructed from six dimensions: completeness, accuracy, comprehensibility, effectiveness, compliance, and security, with quantifiable interpretability levels. The evaluation system should include specific quantitative criteria, such as completeness requirements mandating a log coverage rate of 100%, and accuracy requirements requiring consistency between interpreted results and the actual model logic to exceed 95%.

## 4.3 Legal Liability Dimension

### 4.3.1 Definition of Responsible Entities

To define legal liabilities regarding algorithmic explainability in AI auditing, it is essential to first clarify the boundaries and scope of responsible entities. In traditional auditing scenarios, accountability primarily focuses on direct participants such as auditing institutions and auditors. However, with the introduction of AI auditing algorithms, the range of responsible parties has become more diversified, encompassing algorithm developers, auditing institutions, audited entities, and algorithm users [7]. This diversification stems from the multi-stage nature of algorithm system development, deployment, and application processes—each phase impacts the realization of algorithmic explainability and consequently influences legal liability obligations.

The determination of liability subjects requires integration of algorithmic explainability's technical characteristics with legal regulatory requirements. Technically, the black-box nature of algorithms poses challenges in liability tracing, where the extent of explainability directly defines accountability boundaries: developers employing industry-standard explainable technologies face relatively lighter liabilities, while non-compliant implementations result in significantly heightened fault severity. Legally, current international regulations lack explicit provisions on AI audit accountability. Algorithmic systems should be classified as "electronic audit tools" under applicable laws, adhering to the principle of "whoever develops bears responsibility, whoever uses requires oversight," with specific liability delineation based on operational contexts.

The distinct roles of different responsible parties in algorithmic explainability determine their respective liability classifications. Algorithm developers primarily bear technical responsibilities, ensuring algorithm designs meet explainability standards while providing essential technical documentation and explanatory tools. As the implementing entities of AI audits, auditing institutions must oversee algorithm selection, deployment, and application processes to guarantee decision-making transparency and regulatory compliance. Audited organizations are required to collaborate with auditing bodies by supplying relevant algorithmic data and information, ensuring traceability of algorithmic execution environments. This multi-stakeholder accountability framework requires balancing technical feasibility with legal fairness, preventing ineffective enforcement of legal liabilities due to ambiguous responsibility attribution.

### 4.3.2 Principle of Responsibility Allocation

Under the AI audit framework, the delineation of explainability and legal liabilities must adhere to the principle of technical compatibility, ensuring that accountability assignments align with the technical implementation level of explainability in English algorithms. Current AI audit algorithms predominantly rely on neural networks and deep learning models, whose black-box characteristics significantly increase interpretability challenges. While heuristic coarse-grained algorithms and other training acceleration frameworks can optimize model efficiency, they must still be tailored to meet specific audit scenario requirements.

Responsibility allocation must align with audit supervision logic within corporate governance frameworks to ensure accountable entities comply with auditing standards. The core objective of auditing is to verify financial data authenticity through independence validation, while AI as a technological tool requires responsibility distribution embedded within the enterprise governance ecosystem. Specifically, enterprises as audit entrusting parties must oversee the selection and application of AI auditing tools; audit institutions as executors should conduct manual verification of algorithmic outcomes; and algorithm developers must ensure model explainability. Clarifying these three-party accountability boundaries establishes a responsibility chain comprising "technical safeguards by developers – professional validation by audit institutions – oversight by corporate management."

### 4.3.3 Principle of Burden of Proof

Due to the unique nature of the auditing industry, victims often struggle to obtain critical evidence after reporting infringements of their legitimate rights. Therefore, when loss or infringement incidents occur, the principle of reversed burden of proof is applied alongside the presumption of fault. Responsible parties must demonstrate that the algorithm meets interpretability standards, compliance obligations have been fulfilled,

and damages resulted from third-party actions or force majeure events. Professional institutions must issue appraisal reports to reconfirm algorithmic interpretability and decision-making logic. This process determines whether to overturn the presumption of fault, thereby maximizing protection of stakeholders' lawful rights.

## 5. Conclusion

This study conducts a framework analysis on algorithmic interpretability and legal liability delineation in AI auditing contexts based on theoretical research, yielding the following conclusions: First, insufficient algorithmic interpretability remains a critical bottleneck constraining the standardized development of AI auditing. Key challenges include complex algorithmic technology development, prominent conflicts between interpretability and audit efficiency, and inconsistent standards. These issues make audit conclusions difficult to verify and legal liability determination challenging. Second, legal liability definition for algorithmic interpretability exhibits intrinsic interdependence. Algorithmic interpretability serves as the prerequisite and foundation for legal framework establishment, while refined liability determination rules can drive improvements in interpretability. Third, the proposed legal liability framework effectively addresses algorithmic interpretability deficiencies and ambiguous liability boundaries by clarifying stakeholders' responsibilities and obligations, thereby providing theoretical support and practical guidance for compliant AI auditing development.

This study still has several limitations: Firstly, the exploration of algorithmic interpretability implementation pathways remains at a macro level without in-depth analysis of technical specifics. Secondly, the constructed legal framework lacks practical validation, requiring further optimization for operational feasibility. Thirdly, case selection remains insufficient, with limited representation of current AI audit practices across enterprises, and the generalizability of research conclusions requires verification. Addressing existing challenges in AI auditing, priority should be given to enhancing technical interpretability standards. Aligning with AI technological advancements, research institutions should develop more audit-aligned interpretability technologies and methodologies while strengthening university collaborations to overcome interpretability bottlenecks. Regarding legal liability frameworks, a comprehensive responsibility allocation system must be established to make macro-level accountability assessments, incorporating entry thresholds and verification mechanisms. Simultaneously, case studies spanning large, medium, and small enterprises should be selected to conduct tier-specific framework validation, thereby improving operational applicability.

## References

[1] Xue, L. D., & Qian, Y. [1] Xue, L. D., & Qian, Y. (2025). The realistic dilemmas, logical framework, and implementation paths of algorithmic auditing. Finance and Accounting Communication, (1), 8–16.

[2] Birhane, A., Steed, R., Ojewale, V., Vecchione, B., & Raji, I. D. [2] Birhane, A., Steed, R., Ojewale, V., Vecchione, B., & Raji, I. D. (2024). AI auditing: The broken bus on the road to AI accountability. arXiv:2401.14462 [cs.CY].

[3] Onwubuariri, E. R., Adelakun, B. O., Olaiya, O. P., & Ziorklui, J. E. K. [3] Onwubuariri, E. R., Adelakun, B. O., Olaiya, O. P., & Ziorklui, J. E. K. (2024). AI-driven risk assessment: Revolutionizing audit planning and execution. Finance & Accounting Research Journal, 6(6), 1069–1090.

[4] Wang, H. Y. [4] Wang, H. Y. (2024). The value of algorithmic explainability and its legalization path. Chongqing Social Sciences, (1), 120–135.

[5] The Lancet Digital Health. (2022, May). Holding artificial intelligence to account [Editorial]. The Lancet Digital Health, 4(5), e290.

[6] Mökander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. Digital Society, 2(49), 1–32.

[7] Ayankoya, M. B. [7] Ayankoya, M. B. (2025). Explainable AI in data-driven finance: Balancing algorithmic transparency with operational optimization demands. International Journal of Advance Research Publication and Reviews, 2(6), 125–149.

[8] Deloitte. (2025, January 13). AI reshapes auditing: Exploring the path of industry transformation driven by artificial intelligence.

[9] Lei, G., & Sun, G. M. [9] Lei, G., & Sun, G. M. (2025). From value positioning to institutional construction: An analysis of the path of algorithmic auditing system in the digital age. Gansu Theory Research, (3), 75–84.

[10] Jin, L. J. [10] Jin, L. J. (2025). The unexplainability of generative AI and its legal response. Legal Research, (2), 42–53.

[11] Kumar, N., & Aeron, A.[11] Kumar, N., & Aeron, A. (2025). Ethical AI framework and auditing: A comprehensive overview. International Journal of Sustainable Studies, Technologies, and Assessments, 1(2), 1–13.

[12] van Kolfschooten, H., & van Oirschot, J. (2024). The EU Artificial Intelligence Act (2024): Implications for healthcare. Health Policy, 149, 105152.

[13] Cen, S. H., & Alur, R. [13] Cen, S. H., & Alur, R. (2024). From transparency to accountability and back: A discussion of access and evidence in AI auditing. arXiv:2410.04772 [cs.CY].

[14] O'Neil, C., Sargeant, H., & Appel, J. (2024). Explainable fairness in regulatory algorithmic auditing. West Virginia Law Review, 127, 79–133.

[15] The Lancet Digital Health. (2022, May). Holding artificial intelligence to account [Editorial]. The Lancet Digital Health, 4(5), e290.

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgment

This paper is an output of the science project.

## Open Access