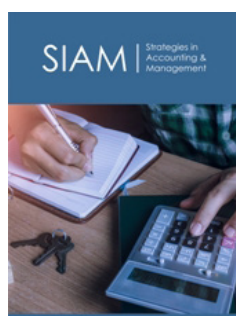


Explainable AI for Risk Management and Auditing

ISSN: 2770-6648

**Mohd Nadeem* and Ankit Singh**

Department of Computer Science and Engineering, Shri Ramswaroop Memorial University, Barabanki, India

Abstract

Artificial Intelligence (AI) is rapidly transforming the way organizations manage risks and conduct audits. Modern businesses generate vast amounts of financial, operational, and transactional data that exceed the capabilities of traditional auditing approaches. AI-driven systems can analyze these large datasets, identify hidden patterns, detect anomalies, and support timely decision-making. However, many advanced AI models operate as “black boxes,” making it difficult for auditors, managers, regulators, and stakeholders to understand how decisions are reached. This lack of transparency can reduce trust and create challenges in regulatory compliance and accountability. Explainable Artificial Intelligence (XAI) has emerged as a promising solution by providing clear and interpretable explanations for AI-generated outcomes. This mini review explores the growing role of XAI in risk management and auditing, examines key explainability techniques, discusses practical applications, and highlights current challenges and future opportunities. The review suggests that integrating XAI into auditing and risk assessment processes can strengthen organizational governance, improve decision quality, and enhance stakeholder confidence in AI-enabled systems.

Keywords: Explainable artificial intelligence; Auditing; Risk management; Transparency; Fraud detection

***Corresponding author:** Mohd Nadeem, Department of Computer Science and Engineering, Shri Ramswaroop Memorial University, Barabanki, India

Submission: 📅 June 08, 2026**Published:** 📅 July 08, 2026

Volume 6 - Issue 2

How to cite this article: Mohd Nadeem* and Ankit Singh. Explainable AI for Risk Management and Auditing. *Strategies in Accounting & Management*, 6(2), SIAM. 000635. 2026. DOI: [10.31031/SIAM.2026.06.000635](https://doi.org/10.31031/SIAM.2026.06.000635)

Copyright@ Mohd Nadeem, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Organizations today operate in increasingly complex and data-rich environments. Financial transactions occur in real time, regulatory requirements continue to expand, and businesses face a wide range of risks, including financial, operational, cybersecurity, and compliance-related threats [1,2]. To address these challenges, many organizations are adopting Artificial Intelligence (AI) technologies to support auditing activities, monitor risks, and improve decision-making processes [3]. AI systems have demonstrated remarkable capabilities in identifying unusual transactions, predicting potential risks, and uncovering patterns that may remain unnoticed through traditional analytical methods [4]. As a result, AI has become an important tool for auditors, financial analysts, and risk managers seeking greater efficiency and accuracy. Despite these advantages, many advanced AI models-particularly deep learning systems-provide little insight into how they generate predictions or recommendations [5]. This lack of transparency creates concerns regarding accountability, fairness, and regulatory compliance [6]. Auditors and managers often require clear explanations before relying on AI-generated decisions, especially when those decisions influence financial reporting, risk assessments, or strategic actions. Explainable Artificial Intelligence (XAI) seeks to bridge this gap by making AI systems more transparent and understandable [7]. Rather than simply providing predictions, XAI helps users understand why a decision was made, which factors influenced the outcome, and how different variables contributed to the final result [8,9]. Consequently, XAI is becoming an essential component of modern risk management and auditing practices.

The evolution of AI in risk management and auditing

Auditing has evolved significantly over the past few decades. Traditional auditing approaches primarily relied on manual inspections, document reviews, and sampling techniques. While these methods remain valuable, they often struggle to keep pace with the volume and complexity of modern business data [10]. The introduction of computer-assisted auditing tools marked the first major technological advancement, allowing auditors to process larger datasets more efficiently. Subsequently, machine learning algorithms enabled organizations to automate anomaly detection, fraud identification, and risk assessments. More recently, deep learning and advanced analytics have further improved predictive capabilities [11,12]. However, as predictive performance increased, model transparency often decreased. This challenge has led to growing interest in explainable AI approaches that combine analytical power with interpretability [13]. Today, organizations are moving toward trustworthy AI ecosystems where predictive accuracy, transparency, accountability, and governance are considered equally important (Figure 1).

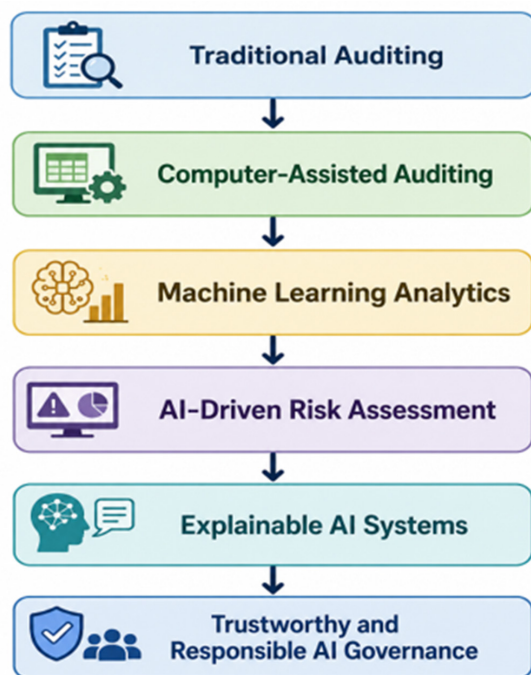


Figure 1: Evolution of auditing technologies.

Explainable artificial intelligence

Explainable Artificial Intelligence refers to a collection of methods and techniques designed to make AI systems more understandable to human users [14]. The primary objective of XAI is not only to generate accurate predictions but also to provide meaningful explanations that allow users to understand and trust those predictions. In auditing and risk management environments, explainability helps answer critical questions:

- A. Why was a transaction classified as suspicious?
- B. Which factors contributed most to a risk score?
- C. How reliable is the model’s prediction?
- D. What actions could change the outcome?

By addressing these questions, XAI enhances transparency and supports more informed decision-making.

Intrinsically Explainable Models: Some AI models are naturally interpretable because their decision-making processes can be easily understood. Examples include:

1. Decision Trees
2. Rule-Based Systems
3. Linear Regression Models
4. Logistic Regression Models

These approaches provide straightforward explanations but may sometimes sacrifice predictive performance when dealing with highly complex datasets. **Post-Hoc Explanation Techniques:** For more complex models such as neural networks, specialized explanation methods are applied after predictions are generated. Common techniques include:

1. SHAP (Shapley Additive Explanations).
2. LIME (Local Interpretable Model-Agnostic Explanations).
3. Feature Importance Analysis.
4. Partial Dependence Plots.
5. Counterfactual Explanations.

These methods help users understand the behavior of otherwise opaque AI systems (Table 1).

Table 1: Common explainable AI techniques in auditing and risk management.

Technique	Type	Main Purpose	Example Application
Decision Trees	Intrinsic	Transparent classification	Risk categorization
Rule-Based Models	Intrinsic	Compliance evaluation	Internal controls assessment
SHAP	Post-hoc	Feature contribution analysis	Credit risk assessment
LIME	Post-hoc	Local prediction explanation	Fraud investigation
PDP	Post-hoc	Variable impact visualization	Financial forecasting
Counterfactual Analysis	Post-hoc	Alternative outcome exploration	Audit decision support

Applications of XAI in risk management

The growing adoption of AI has created new opportunities for improving risk management practices. However, effective risk management requires not only accurate predictions but also clear reasoning behind those predictions.

Financial risk assessment: Financial institutions increasingly use AI models to evaluate creditworthiness, predict loan defaults, assess liquidity risks, and forecast market fluctuations. Explainable AI helps analysts understand which variables influence risk scores and how those variables interact within predictive models. This transparency enables better decision-making while ensuring compliance with regulatory expectations [15].

Fraud detection and prevention: Fraud remains a major concern across industries. Machine learning algorithms can rapidly identify suspicious transactions and unusual behavioral patterns. However, auditors often need evidence explaining why a transaction was flagged [16].

XAI provides these explanations by highlighting influential features, transaction characteristics, and behavioral indicators associated with fraudulent activities. This capability significantly improves the effectiveness of fraud investigations [17].

Enterprise risk management: Organizations increasingly integrate AI into Enterprise Risk Management (ERM) frameworks. XAI supports strategic risk evaluation by making risk predictions easier to interpret and communicate [18].

Managers can better understand emerging threats, prioritize mitigation strategies, and justify decisions to stakeholders and regulatory authorities [19] (Figure 2).

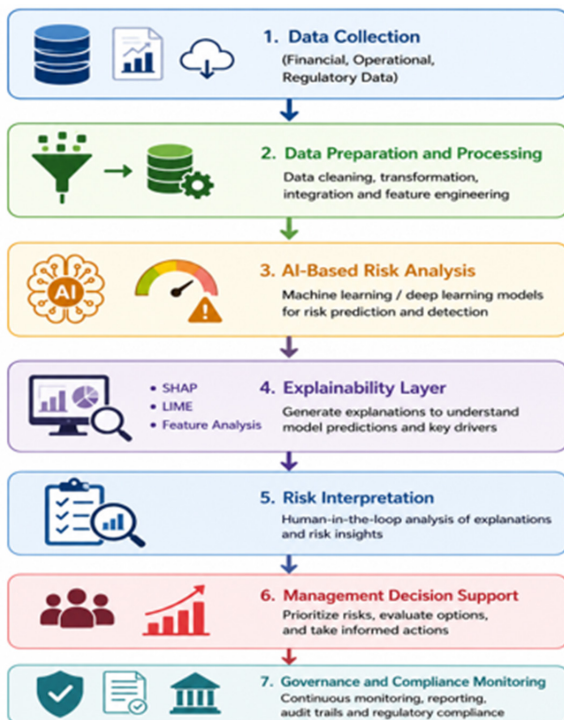


Figure 2: XAI enabled risk management framework.

The role of explainable AI in auditing

Auditing is fundamentally built on transparency, evidence, and accountability. As AI becomes more deeply embedded in audit processes, explainability becomes increasingly important. Modern AI-powered auditing systems can:

1. Monitor transactions continuously.
2. Detect unusual accounting activities.
3. Evaluate internal controls.
4. Identify compliance violations.
5. Assess financial reporting risks.

While these capabilities improve efficiency, auditors must still be able to justify their conclusions. Explainable AI enables auditors to understand how AI systems reach their recommendations, ensuring that professional judgment remains central to the audit process. Moreover, XAI strengthens communication between auditors, management teams, regulators, and external stakeholders by providing clear and understandable explanations [20] (Table 2).

Table 2: Benefits of explainable ai in auditing.

Benefit	Organizational Value
Transparency	Improves understanding of AI decisions
Trust	Increases stakeholder confidence
Accountability	Enhances audit traceability
Compliance	Supports regulatory requirements
Validation	Enables model verification
Better Decisions	Improves audit quality and reliability

Governance, ethics, and regulatory considerations

As AI adoption expands, governments and regulatory bodies are placing greater emphasis on transparency, accountability, and ethical AI practices. Organizations must ensure that AI systems operate fairly and can justify their decisions when challenged [21,22]. Explainable AI contributes significantly to responsible AI governance by:

1. Supporting regulatory compliance.
2. Identifying potential biases.
3. Enhancing model accountability.
4. Improving documentation and reporting.
5. Facilitating independent audits of AI systems.

For auditors and risk managers, explainability is becoming not merely a technical feature but a governance requirement [23].

Challenges and limitations

Although XAI offers substantial benefits, several challenges continue to limit widespread adoption.

Accuracy versus interpretability: Highly interpretable models are often easier to understand but may not achieve the same predictive performance as complex deep learning models.

Computational complexity: Generating explanations can increase computational costs, particularly for large-scale enterprise systems [24].

Lack of standardized frameworks: Organizations currently employ different explanation techniques, making comparisons and standardization difficult.

Regulatory uncertainty: AI-related regulations continue to evolve globally, creating uncertainty regarding future compliance requirements.

Human interpretation challenges: Even when explanations are available, users may misunderstand or misinterpret the information provided (Table 3).

Table 3: Challenges and potential solutions.

Challenge	Impact	Possible Solution
Complex AI Models	Reduced transparency	Hybrid explainable architectures
High Computational Costs	Increased processing time	Efficient explanation algorithms
Bias and Fairness Issues	Unfair outcomes	Bias monitoring frameworks
Regulatory Changes	Compliance uncertainty	Adaptive governance models
Misinterpretation of Explanations	Poor decision-making	User training and education

Future research opportunities

The field of Explainable AI is evolving rapidly, creating numerous opportunities for future research and innovation.

Explainable deep learning: Researchers are developing new approaches to make deep neural networks more transparent without sacrificing performance.

Responsible and Ethical AI auditing: Future audit frameworks are expected to incorporate ethical assessments alongside traditional financial evaluations.

Federated explainable AI: Combining privacy-preserving learning with explainability could enable secure and transparent

collaboration across organizations.

Explainable generative AI: As large language models and generative AI systems become more prevalent, explainability will be essential for validating generated outputs.

ESG and sustainability auditing: Organizations increasingly require transparent AI systems to evaluate environmental, social, and governance performance indicators.

Human-centered AI governance: Future systems will likely emphasize collaboration between human expertise and AI capabilities, ensuring that decision-making remains accountable and trustworthy (Figure 3).

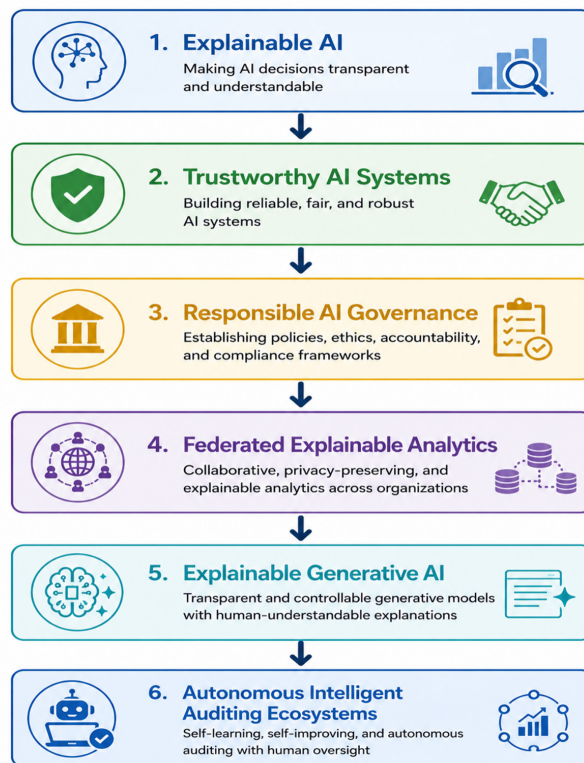


Figure 3: Future roadmap for explainable ai in auditing.

Conclusion

Explainable Artificial Intelligence represents a critical advancement in the evolution of modern risk management and auditing. While AI technologies offer unprecedented capabilities for data analysis, prediction, and automation, their effectiveness ultimately depends on user trust and transparency. Explainable AI addresses these concerns by providing meaningful insights into how decisions are generated, thereby enabling auditors, managers, and regulators to understand and validate AI-driven outcomes. The integration of XAI into auditing and risk management enhances accountability, strengthens governance frameworks, improves regulatory compliance, and supports better organizational decision-making. Although challenges related to complexity, standardization, and evolving regulations remain, ongoing research continues to improve the interpretability and reliability of AI systems. As organizations increasingly adopt intelligent technologies, explainability will become a fundamental requirement rather than an optional feature. The future of auditing and risk management is likely to be characterized by AI systems that are not only accurate and efficient but also transparent, trustworthy, and aligned with organizational and societal expectations.

References

- Kirmani S, Raghavan P (2013) Scalable parallel graph partitioning. In SC 13: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, pp. 1-10.
- Kirmani F, Lane BJ, Rose JR (2019) Exploring machine learning techniques to improve peptide identification. In 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 66-71.
- Kirmani S, Park J, Raghavan P (2017) An embedded sectioning scheme for multiprocessor topology-aware mapping of irregular applications. *Int J High Perform Comput Appl* 31(1): 91-103.
- S. Kirmani S, Sun H, Raghavan P (2018) A scalability and sensitivity study of parallel geometric algorithms for graph partitioning. In 2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), pp. 420-427.
- Kirmani F, Lane B, Rose J (2025) Identifying proteotypic peptides via deep learning. In Proceedings of the 11th International Conference on Bioinformatics Research and Applications, pp. 42-47.
- Tyler J, Pastor J, Huhns MN, Kirmani S, Du H (2013) Exposing, formalizing and reasoning over the latent semantics of tags in multimodal data sources. *Appl Ontol* 8(2): 95-130.
- Kirmani S, Shankar M (2022) Generating keywords by associative context with input words. Google Patents.
- Mishra A, Kirmani S, Madduri K (2020) Fast spectral graph layout on multicore platforms.
- Kirmani S, Madduri K (2018) Spectral graph drawing: Building blocks and performance analysis. In 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 269-277.
- Alyami H, Nadeem M, Alosaimi W, Alharbi A, Kumar R, et al. (2022) Analyzing the data of software security life-span: Quantum computing era. *Intell Autom Soft Comput* 31(2): 707-716.
- Ahmad M, Al-Amri J (2022) Healthcare device security assessment through computational methodology. *Comput Syst Sci Eng* 41(2): 811-828.
- Alyami H (2021) The evaluation of software security through quantum computing techniques: A durability perspective. *Appl Sci* 11(24).
- Samuel O, Javaid N, Alghamdi TA, Kumar N (2022) Towards sustainable smart cities: A secure and scalable trading system for residential homes using blockchain and artificial intelligence. *Sustain Cities Soc* 76: 103371.
- Attaallah A, Khatri S, Nadeem M, Ansar SA, Pandey AK, et al. (2021) Prediction of COVID-19 pandemic spread in Kingdom of Saudi Arabia. *Comput Syst Sci Eng* 37(3): 313-329.
- Alharbi A, Alosaimi W, Alyami H, Nadeem M, Faizan M, et al. (2021) Managing software security risks through an integrated computational method. *Intell Autom Soft Comput* 28(1): 179.
- Alharbi A, Alosaimi W, Nadeem M, Alyami H, Alouffi B, et al. (2025) Novel 59-layer dense inception network for robust deepfake identification. *Sci Rep* 15(1): 24159.
- Hakami A, Mohammed Badedi M, Elsiddig M, Nadeem M, Altherwi N, et al. (2021) Clinical characteristics and early outcomes of hospitalized COVID-19 patients with end-stage kidney disease in Saudi Arabia. *Int J Gen Med* 14: 4837-4845.
- Nadeem M, Pathak PC, Ahmad M, Farooqui NA (2024) Identification of security factors in cloud computing: Defence security perspective. in *Computational Intelligence Applications in Cyber Security*, CRC Press, pp. 78-99.
- Nadeem M (2024) Deep learning approach for classifying DDoS attack traffic in SDN environments. *J Inf Secur Cybercrimes Res* 7(2): 109-126.
- Nadeem M, Ahmad M, Ahmad M, Pathak PC, Gupta S, et al. (2023) Evaluating the factors of CGTMSE scheme in bank by using fuzzy AHP. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) 6: 56-61.
- Nadeem M (2024) Analyze quantum security in software design using fuzzy-AHP. *Int J Inf Technol*.
- Alosaimi W, Alharbi A, Alyami H, Alouffi B, Almulihi A, et al. (2024) Analyzing the impact of quantum computing on IoT security using computational based data analytics techniques. *AIMS Math* 9(3): 7017-7039.
- Allassery F, Alzahrani A, Khan AI, Khan A, Nadeem M, et al. (2022) Quantitative evaluation of mental-health in type-2 diabetes patients through computational model. *Intell Autom Soft Comput* 32(3): 1701-1715.
- Alharbi A (2024) Selection of data analytic techniques by using fuzzy AHP TOPSIS from a healthcare perspective. *BMC Med Inform Decis Mak* 24(1): 240.