Blockchain + AI for Transparent and Auditable AI Models

Tejal Pegwar*, Ruhina Siddiqui

Department of Computer Science, Dr. D.Y. Patil Arts, Commerce & Science College, Pimpri Pune-18, Maharashtra, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP013

Received: 26 June 2025; Accepted: 30 June 2025; Published: 23 October 2025

Abstract: As AI becomes deeply entrenched in mission-critical domains like healthcare, finance, and government, the need for reliable, explainable, and ethically compliant AI systems has increased immensely. Most traditional AI systems exist as opaque "black boxes" wherein it is not possible to see how decisions are being made or to verify compliance with rules and ethical requirements. This transparency issue makes it challenging to hold AI systems accountable and develop confidence in the outcomes produced by them. This work presents a new framework that marries the advantages of blockchain technology with explainable AI to produce transparent and auditable AI systems. The essential properties of blockchain decentralization, immutability, and automation of smart contracts are utilized to have tamper-proof records of the whole AI life cycle. This entails data gathering, preprocessing, and training of models, updates, and inference events. These logs create an unalterable audit trail that allows regulators, users, and stakeholders to confirm the integrity and fairness of the AI models at any given moment. Moreover, the framework incorporates explainable AI methods to produce human-interpretable explanations of model outputs. This not only enhances transparency but also enables stakeholders to determine whether AI judgments are reasonable and unbiased. We provide a prototype implementation of this framework and compare its performance in a real-world case study in the healthcare industry. The findings show that the integrated system effectively strengthens traceability, establishes trust, and facilitates regulatory compliance without any decline in the performance of AI models. In summary, this study demonstrates how the integration of blockchain and AI closes essential gaps in transparency and accountability, paving the way for the responsible and ethical use of AI. The framework outlined provides a pragmatic way forward for companies wishing to implement AI technologies without diminishing public trust and fulfilling legal requirements.

Keywords: Blockchain, Artificial Intelligence, Decentralization, Immutability, Transparency, AI Decision Traceability.

I. Introduction

Artificial Intelligence (AI) is fast changing decision-making in key industries like finance, healthcare, autonomous systems, and public administration. Although AI models are highly efficient and predictive, they tend to be opaque "black boxes" and are not easily traceable for determining how individual decisions are arrived at. Such lack of transparency poses serious ethical, legal, and operational issues especially when AI technologies are applied in sensitive or regulated domains. With increasing demands for explainable, fair, and accountable AI, come the demands for frameworks that can furnish verifiable evidence of model behavior, data usage, and decision logic. Blockchain technology, with its own attributes of immutability, decentralization, and openness, offers a strong solution to this problem. Utilizing blockchain, one can establish a tamper-evident and secure audit trail of the lifecycle of the AI model such as data provenance, training parameters, model updates, and inference outputs. Additionally, the use of smart contracts allows for automated compliance enforcement of rules, data access regulations, and governance structures. This work suggests hybrid architecture based on blockchain and AI, which would allow transparent and auditable AI systems. The architecture traces important aspects of AI development and deployment to a distributed ledger and incorporates tools for explainability to capture justifications for model decisions. The aim is to provide stakeholders developers, regulators, and users with a capability to inspect, verify, and trust AI-driven processes. By this interdisciplinary endeavor, the research makes contributions to the still developing area of trustworthy AI and is in accordance with international efforts on responsible and ethical deployment of AI.

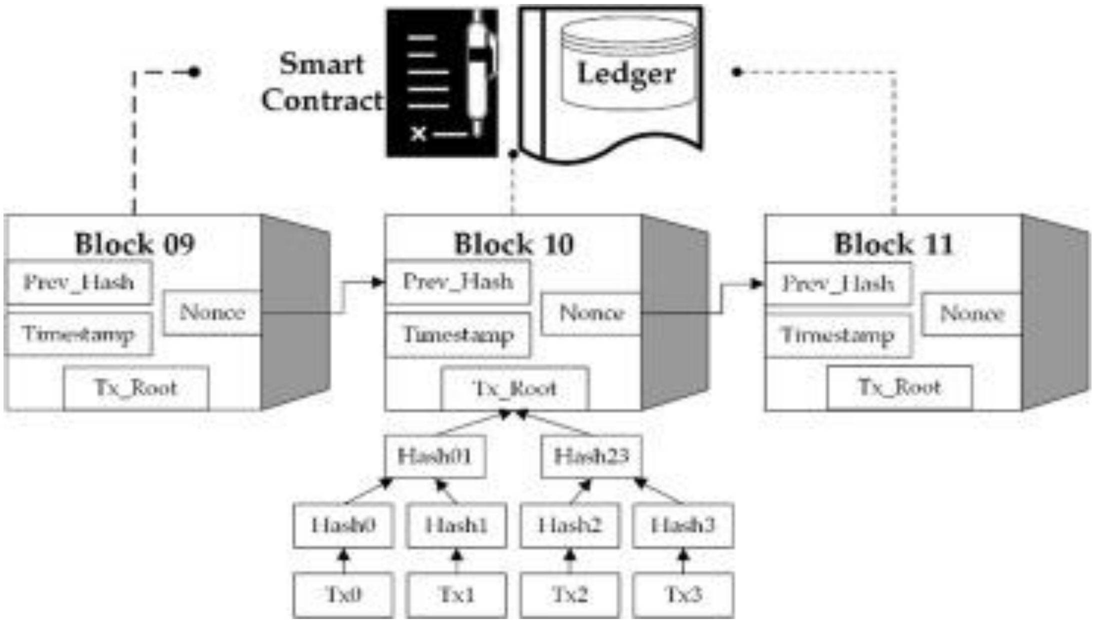
II. Literature Review

AI in Financial Transactions

Artificial Intelligence (AI) is a revolutionary technology that has gained a foothold in the finance industry, especially with regards to identifying suspicious transactions, risk management, and predictive analytics. Artificial intelligence systems employ machine learning programs to process vast volumes of transaction data, identifying patterns and anomalies that may indicate fraudulent behavior. As per different studies, AI-based fraud detection systems have the ability to improve the accuracy of flagging suspicious behavior by constantly learning and evolving as per new fraud methods. This ability helps the financial institutions minimize false positives while responding quickly to potential threats. Further, AI has become inseparable as a component of risk management in financial activities. Through historical data analysis, AI models can forecast potential risks, providing valuable insights enabling financial institutions to make better decisions. In predictive analytics, AI allows for the creation of customized financial services through predicting customer behavior, increasing client retention, and optimizing product recommendation. These innovations render AI an essential element in the process of modernizing financial transactions, which results in improved efficiency of operations and customer satisfaction.

Blockchain Technology Overview

Blockchain technology essentially functions as a shared, unchangeable record book that streamlines how transactions are logged and assets are tracked within a business network. An asset can be anything from physical items like a house, car, cash, or land, to intangible goods such as intellectual property, patents, copyrights, or brand identity. Practically any item of value can be monitored and exchanged on a blockchain network, which ultimately reduces risks and lowers costs for all participants. The name "blockchain" itself comes from how transactional data is stored. It's structured as a series of blocks connected together, forming a chain. Each block contains a collection of transactions, often organized in a Merkle tree. This structure allows for the cryptographic hashes of these transactions to be stored in the block's header, making it quick and easy to verify individual transactions within that block. Beyond the transactions, each block also includes a timestamp (marking its creation time), the hash of the preceding block, a unique block ID or number, and a proof generated by the consensus algorithm (often called a "nonce"). Once a block is formed, it's integrated into the distributed ledger. This process isn't random; it strictly follows consensus rules that all participants in the network have previously agreed upon. For a new block to be permanently recorded, it must first be validated and approved by all network participants. To fully grasp how blockchain works and to understand any proposed frameworks built upon it, it's crucial to be familiar with its four core components: the distributed shared ledger, smart contracts, permissions, and consensus mechanisms.



Integrating AI and Blockchain: A Powerful Synergy

The blend of Artificial Intelligence (AI) and Blockchain technology has recently emerged as a compelling force, especially when it comes to boosting the security and transparency of financial transactions. Current research strongly indicates that AI algorithms can play a crucial role in automating and enhancing the verification processes within Blockchain networks.

By leveraging AI's robust data analysis strengths, Blockchain systems can significantly improve transaction validation, leading to more efficient detection of fraudulent activities. For instance, convolutional neural networks (CNNs) have proven invaluable in Blockchain based security systems for tasks like image recognition, drastically improving how anomalies in transaction data are identified. Furthermore, reinforcement learning models are being used to refine the decision-making in AI-driven smart contracts, allowing these systems to adapt dynamically to new transaction conditions without human intervention. We've also seen decision trees used to verify transaction integrity before data is entered into the Blockchain ledger; ensuring only legitimate information is recorded. This highlights the intricate and sophisticated ways these technologies are being combined.

Let's dive into some key advantages of this integration:

Accelerated Blockchain Verification: AI algorithms can dramatically speed up the verification process on a Blockchain. They achieve this by rapidly analyzing and validating large volumes of transactions. This not only reduces the time needed to reach consensus but also makes fraud detection within the Blockchain network far more accurate.

Improved Data Handling: AI helps Blockchain systems scale up by processing and analyzing transaction data in real-time. This enables these networks to manage much larger transaction volumes without compromising security or speed.

Intelligent Smart Contracts: AI-powered smart contracts offer another exciting area of integration. These contracts can self-execute based on predefined conditions and, with AI algorithms embedded, can also assess the validity of each transaction. This combination significantly enhances the efficiency and trustworthiness of automated financial agreements, effectively removing the need for intermediaries.

Ultimately, pairing AI's impressive predictive capabilities with Blockchain's inherently secure framework creates a highly robust system for financial transactions one that's both transparent and secure. This integration opens up promising avenues for future research and practical applications, especially in fields demanding high levels of security, such as financial services and supply chain management.

III. Research Methodology

System Design

The system is built on three interconnected layers to ensure transparency, security, and intelligent analysis. At the core lies the Blockchain Layer, which uses a permission blockchain platform like Hyper ledger Fabric to control access and maintain the integrity of the data. This layer records metadata related to AI training datasets such as data sources, preprocessing methods, and version history as well as details of the AI models themselves, including their structure and parameters. It also logs every interaction with the AI system, including inputs, outputs, and decisions, to create a fully auditable and tamper-proof record. On top of this sits the AI Layer, which utilizes the recorded data for training and validating AI models. This layer applies various AI techniques to uncover patterns, detect anomalies, and flag potential biases in the system. To enhance transparency, it incorporates explainable AI methods, offering clear reasoning behind AI-generated outcomes. Furthermore, it includes AI-driven anomaly detection to help spot irregular or potentially dishonest activities. Finally, the Integration Layer ensures smooth and secure interaction between the Blockchain and AI components. It establishes encrypted communication channels and uses smart contracts to automate essential operations like data validation, access control, and reporting. This layer also provides application programming interfaces (APIs) to facilitate user interaction with the system, making it easier to query and manage data.

Steps for Developing a Prototype

Identify the Use Case: Begin by selecting a specific area of application such as financial auditing, supply chain management, or another relevant domain to effectively showcase what the system can do.

Choose the Right Technologies: Choose the appropriate tools and technologies tailored to each layer of the system. This includes selecting a blockchain platform like Hyperledger Fabric or Corda, AI frameworks such as TensorFlow or PyTorch, and programming languages like Python or Go, depending on the project’s needs.

Build the Core Components: Develop the foundational parts of the system. This involves setting up the blockchain infrastructure (like smart contracts and data handling), creating and training AI models, and developing integration modules that allow seamless communication between components.

Integrate All Parts: Ensure the blockchain and AI layers work together smoothly. Establish secure channels for data transfer and manage the overall workflow through proper orchestration and APIs.

Test the System: Carry out extensive testing to assess the system’s functionality, security, performance, and level of transparency. This helps ensure the prototype meets both technical and user requirements.

Refine and Improve: Use the feedback from testing and potential users to make improvements. Continue refining the prototype until it performs reliably and effectively.

Evaluation Metrics

To evaluate the effectiveness of a Blockchain-AI integrated system, several key metrics can be considered. Transparency plays a crucial role, focusing on how well the system can track the origin of data; the AI model’s training process, and the reasoning behind its decisions. Auditability refers to how easily the system’s operations can be verified and whether any unauthorized changes or tampering can be detected. Performance involves analyzing how efficiently the system runs, including response times, processing speed, and the amount of computational resources it uses. Security looks at the system’s ability to withstand potential threats and protect against vulnerabilities. Lastly, scalability assesses whether the system can maintain its performance as data volume and the number of transactions increase.

A practical example of such a system can be found in supply chain traceability. In this scenario, blockchain technology is used to document every stage in the lifecycle of a product, from its origin and manufacturing to its transport and delivery. AI models then process this blockchain data to identify risks such as delays, supply disruptions, or counterfeit items. The result is a supply chain system that is not only efficient and predictive but also transparent and verifiable helping to build greater trust among all involved parties.

Verification Mechanism

Understanding the Verification Medium of Zero-Knowledge Attestations (ZKPs) Zero-Knowledge Attestations (ZKPs) are advanced cryptographic ways that allow one party, known as the prover, to move another party, the verifier, that a certain claim is true without telling any fresh information beyond the fact that the claim itself is valid. This system is especially important when dealing with sensitive or private data. It can, for illustration, prove that a nonpublic input satisfies certain conditions or that a machine literacy model generated a correct result, all without exposing the data, calculation sense, or the result itself

How ZKP Verification Works: A Conceptual Overview

The Zero-Knowledge Proof process generally unfolds in three crucial stages:

- Proof Generation: The prover creates cryptographic evidence that confirms a calculation has been carried out directly.
- Verification: The verifier evaluates the evidence using public information and a verification key to check its correctness.
- Acceptance: The verifier accepts the result as valid — without ever seeing the original input or the internal way of the calculation, If the verification succeeds.

Algorithm Verifiable Computation using zk- SNARKs

Inputs

- Private input x
- Public affair y = f(x)
- A fine representation of the function f, expressed as a circuit C
- A trusted setup phase that generates cryptographic keys:
- Proving Key (PK)
- Verification Key (VK)

Output

A Boolean result that confirms whether the computation was validated correctly.

Procedure:

Trusted Setup:

$$(PK, VK) \leftarrow \text{Setup}(C)$$

A single setup phase that generates cryptographic keys derived from the structure of the computation circuit.

Proof Generation by the Prover:

$\pi \leftarrow \text{Create Proof}(PK, x, y)$

The prover uses the proving key and private input x and public output y to compute a proof π.

Proof Verification by the Verifier:

$\text{isValid} \leftarrow \text{VerifyProof}(VK, y, \pi)$

The verifier verifies the proof with respect to the verification key and the public information.

IV. Result and Discussion

Return is Valid

If is Valid is true, the calculation is confirmed to be accurate.

Advantages of zk-SNARK-Based Verification

Confidentiality: Ensures that inputs and the underlying computation remain private

Efficiency: Enables rapid verification regardless of the complexity of the computation

Non-Interactivity: Requires just a single proof to be shared—no back-and-forth communication is necessary

Use Cases for Blockchain and AI

Throughout diligence, applying AI to blockchain brings new openings

Healthcare: AI can be used to propel nearly every aspect of healthcare, from bringing treatment perceptivity to light and enabling stoner conditions through to detecting perceptivity from patient data and uncovering patterns. With case data stored on blockchain, similar as electronic health records, associations have the capability to unite in order to enhance care while securing patient confidentiality.

Life Sciences: Blockchain and artificial intelligence in the pharmaceutical sector can give translucency and traceability to the pharmaceutical force chain while significantly perfecting the rate of successful clinical trials. The combination of sophisticated data analysis with a decentralized clinical trial frame allows data integrity, translucency, shadowing of cases, operation of concurrence, and automating trial engagement and data accession.

Financial Services: Blockchain and AI are revolutionizing the financial sector by fostering trust, streamlining interactions between multiple parties, and increasing transaction speed. Take the example of a loan application: borrowers provide permission to access their personal information securely stored on the blockchain. Because the data is reliable and evaluation is handled by automated systems, the approval process becomes quicker leading to faster loan closures and a better overall customer experience.

Supply Chain: By digitalizing a process that is generally paper- grounded, making the data shareable and dependable, and investing intelligence and robotization to perform deals, AI and blockchain are reshaping supply chains in colorful diligence and opening up new possibilities. As an illustration, a manufacturer can cover carbon emigrations' data at product or element position, investing delicacy and intelligence to decarburization enterprise.

The fusion of Blockchain and Artificial Intelligence offers a practical way to overcome major challenges related to transparency, security, and accountability in AI systems. By bringing together blockchain’s secure, tamper-proof records and decentralized control with AI’s ability to analyze data and provide meaningful insights, this framework enables the creation of systems that are both intelligent and easy to audit. The integration of zk-SNARKs adds another layer of protection by allowing secure verification without exposing private information. When applied in real-world areas like healthcare, finance, and supply chains, this combined approach has shown strong results in improving trust, meeting regulatory standards, and preserving AI performance. Overall, the model presents a reliable and scalable path toward building responsible and transparent AI systems for critical applications.

V. Conclusion

The combination of Blockchain and Artificial Intelligence offers a powerful solution to key challenges like transparency, accountability, and trust in AI-based systems especially in critical areas such as healthcare, finance, and supply chain management. By using blockchain’s features like decentralization and immutability alongside AI’s ability to analyze and predict, the framework ensures that every stage of the AI process from data collection and model training to decision-making can be tracked and verified. Integrating explainable AI adds another layer of clarity, allowing users to understand and trust AI-generated results. Additionally, incorporating zk-SNARKs boosts privacy and security by enabling result verification without revealing sensitive data. Real-world examples from multiple industries show that this approach is both effective and practical. Overall, this integrated system supports ethical and secure use of AI, helping organizations build greater trust while meeting regulatory demands.

Reference

1. Blockchain and AI Convergence: Creating Explainable, Auditable, and Immutable Data Ecosystems, Vol. 15 No. 1 (2023): IJCMI-VOL15-NO 1,2,3-2023
2. AI and Blockchain Integration: Enhancing Security and Transparency in Financial Transactions, Daniel Martinez, Lena Magdalena, Novalita Savitri, International Transactions on Artificial Intelligence (ITALIC), Vol.3, No.1, November, 2024, pp.11
3. Blockchain Based Audit Trails for AI Model Training Transparency, December2024, Jerry Cole, https://www.researchgate.net/publication/387526768_Blockchain-Based_Audit_Trails_for_AI_Model_Training_Transparency
4. Blockchain-based framework for supply chain traceability: A case example of textile and clothing industry, Tarun Kumar Agrawal, Vijay Kumar, Rudrajeet Pal, Lichuan Wang, Yan Chen, Engineering Volume, April 2021, 107130
5. https://www.ibm.com/think/topics/blockchain-ai
6. Understanding Zero-Knowledge Proofs: Part 1— Verifiable Computation with zk-SNARKs, Bhaskar Krishnamachari, Jul 14, 2024
7. Blockchain for secure and decentralized artificial intelligence in cyber security: A comprehensive review, Ahmed M. Shamsan Saleh, Blockchain: Research and Applications.