World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3062

BLOCKCHAIN AND LARGE LANGUAGE MODELS: A SURVEY OF INTEGRATION APPROACHES AND OPEN CHALLENGES

Yun Li, ShuRui Xiao*
School of Finance and Economics, Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China.
Corresponding Author: ShuRui Xiao

Abstract: This paper surveys recent research on the integration of blockchain technology and Large Language Models (LLMs), examining how decentralization, immutability, and cryptographic verification can address challenges in AI governance, data sovereignty, and model accountability. We review key integration strategies including blockchain-based data repositories, smart contract-driven governance mechanisms, and decentralized AI marketplaces. Our analysis identifies three primary application domains—cybersecurity, healthcare data sharing, and financial services—where this convergence demonstrates practical value. However, significant barriers remain: current blockchain throughput falls short of high-frequency AI requirements, privacy-preserving techniques introduce substantial computational overhead, and regulatory frameworks have yet to reconcile blockchain immutability with data protection mandates. We conclude by proposing prioritized research directions to bridge these gaps.
Keywords: Blockchain; Large language models; Decentralization; Smart contracts; Federated learning

1 INTRODUCTION

Blockchain technology, with its fundamental principles of decentralization, immutability, and transparency, offers potential solutions to critical challenges in AI governance and deployment. Large Language Models (LLMs), while demonstrating exceptional capabilities in natural language understanding and generation, raise concerns regarding centralized control, data privacy, and accountability. The motivation for integrating these technologies stems from structural limitations in the current AI development paradigm. The rapid advancement of AI has been accompanied by increasing concentration of computational resources and training data among a small number of technology corporations [1,2]. This centralization creates several interrelated challenges: first, it can introduce systematic biases in model outputs when training data lacks diversity or adequate curation; second, it restricts broader participation in AI development, limiting innovation pathways; third, it raises questions about the provenance and integrity of model-generated content. Users who contribute data to centralized platforms often lack transparency regarding how their contributions are utilized and monetized.

Blockchain technology offers potential solutions to these challenges through its inherent properties. The decentralized nature of blockchain provides a foundation for more democratic governance of AI systems, addressing eight major governance challenges identified by researchers concerning decision rights, incentive mechanisms, and accountability in foundation model systems [1].

Blockchain implementations tailored for AI integration have evolved beyond traditional cryptocurrency applications to address the specific needs of AI systems. Recent advancements in consensus mechanisms demonstrate this adaptation, with Proof of Authority (PoA) showing promising results in multi-cloud environments, achieving throughputs of up to 1000 transactions per second [3]. While this represents improvement over earlier Proof of Work implementations (typically 7-15 tps for Bitcoin and Ethereum), the requirements for real-time AI operations remain substantially higher, particularly for applications involving continuous data streaming or high-frequency model updates.

Smart contracts have emerged as a key element of blockchain-AI integration, enabling automated governance, verification, and incentive distribution. Research on cybersecurity applications demonstrates that AI algorithms integrated with blockchain-based smart contracts can detect and mitigate network attacks, achieving a 95% detection rate with a 2% false positive rate, while maintaining transaction latency below 200 milliseconds [4]. These performance metrics indicate that for certain AI applications—particularly those tolerant of sub-second latency—smart contracts can provide automated enforcement of governance policies while maintaining reasonable responsiveness.

LLMs have evolved rapidly in recent years, with models such as GPT-4, PaLM, and LLaMA demonstrating capabilities in natural language processing and content generation. However, this development has primarily occurred within centralized environments controlled by resource-intensive organizations, creating barriers to broader participation and raising concerns about bias, accountability, and equitable access.

Blockchain-based approaches aim to address these centralization challenges through several mechanisms. Decentralized governance-driven architectures have been proposed to distribute decision-making power across stakeholders rather than concentrating it within single organizations [1]. These architectures employ smart contracts to implement transparent governance rules, token-based incentive systems to reward contributions to model training or data provision, and cryptographic verification to ensure the integrity of training processes. While practical implementations remain in

early stages, these approaches represent attempts to create more inclusive and accountable frameworks for LLM development.

2 KEY INTEGRATION STRATEGIES

2.1 Decentralized Data Repositories for AI Training

Decentralized data repositories represent a fundamental integration strategy, addressing limitations in data accessibility and sovereignty for LLM training. These systems leverage blockchain to provide verifiable data provenance, decentralized access control, and transparent transaction records. The Blockchain-Based Knowledge Repository (BBKR) exemplifies this approach through structured data validation processes and scalable transaction mechanisms, demonstrating performance improvements in data retrieval and integrity verification compared to centralized alternatives [5]. Building on similar principles, DataHarbour implements a marketplace model specifically designed for AI training data, where blockchain-based smart contracts mediate data transactions between providers and consumers [6]. This marketplace approach aims to address data inequality by enabling smaller organizations to monetize their data assets and access diverse datasets without relying on centralized aggregators.

These decentralized data repositories address a fundamental challenge in LLM development: access to diverse, high-quality training data. By creating secure, transparent, and incentivized mechanisms for data sharing, they enable a broader range of participants to contribute to AI development while maintaining the privacy and security standards crucial for sensitive data.

2.2 Smart Contracts for Model Governance

Smart contracts provide programmable mechanisms for implementing governance policies in decentralized AI systems. These self-executing agreements can automate key governance functions including access control, contribution verification, reward distribution, and dispute resolution, while maintaining a transparent and immutable record of all governance decisions.

AI marketplaces demonstrate practical applications of smart contract-based governance. AIArena implements an on-chain consensus mechanism where participant contributions—whether providing training data, computational resources, or model improvements—are verified and rewarded according to predefined rules encoded in smart contracts [2]. This approach aims to create transparent incentive structures that encourage quality contributions. PredictChain extends this concept by enabling users to upload datasets, request model training on existing datasets, or query trained models, with all interactions mediated by smart contracts executed across blockchain nodes [7]. By encoding governance rules in verifiable and enforceable smart contracts, these systems aim to establish more accountable frameworks for collaborative AI development.

3 EMERGING APPLICATION DOMAINS

3.1 Decentralized AI Marketplaces

Decentralized AI marketplaces represent an emerging application domain where blockchain infrastructure facilitates peer-to-peer exchange of AI-related resources including training data, computational capacity, and trained models. These marketplaces aim to reduce barriers to AI development by eliminating centralized intermediaries. AIArena, implemented on the Base blockchain testnet, demonstrates this concept through a platform where participants can contribute models and computational resources, with contributions verified and rewarded through on-chain mechanisms [2]. DataHarbour focuses specifically on the data acquisition challenge, creating a marketplace where smaller organizations can access diverse datasets without relying on dominant technology providers [6]. PredictChain extends marketplace functionality to encompass the complete AI workflow, enabling dataset uploads, model training requests, and inference queries, all coordinated through blockchain-based protocols [7]. These platforms collectively illustrate how decentralized architectures can broaden participation in AI development, though questions regarding data quality assurance, computational efficiency, and economic sustainability require further investigation.

3.2 Cybersecurity and Threat Detection

Cybersecurity applications demonstrate synergistic integration of blockchain and AI technologies. In this domain, blockchain provides an immutable audit trail of security events and system states, while AI algorithms analyze these records to detect anomalous patterns indicative of attacks. Research on smart city and Industry 4.0 environments shows that this combination can improve both detection speed and accuracy compared to centralized logging systems, where audit records may be subject to tampering [4]. Similarly, electronic voting systems have explored AI-enhanced anomaly detection within blockchain transaction patterns to identify potential security threats in real-time [8]. The mutual reinforcement in these applications—where blockchain secures the data foundation for AI analysis, and AI protects blockchain networks from sophisticated attacks—illustrates a key advantage of integration. However, the computational overhead of maintaining cryptographic verification alongside real-time AI inference remains a practical consideration for resource-constrained deployments.

3.3 Healthcare Data Sharing and Analysis

Healthcare data sharing represents a domain where blockchain-AI integration addresses critical challenges related to patient privacy, institutional trust, and regulatory compliance. Medical data is typically fragmented across institutions, with data sharing constrained by privacy regulations such as HIPAA in the United States and GDPR in Europe.

Federated Learning integrated with blockchain offers a technical approach to this challenge by enabling collaborative AI model training across multiple healthcare institutions without requiring raw patient data to leave institutional boundaries [9]. In this architecture, each institution trains model components locally on its own data, and blockchain coordinates the aggregation of model updates while maintaining a verifiable record of contributions. This approach aims to balance the utility of large-scale datasets for AI model improvement with the privacy requirements of sensitive medical information. However, practical deployment faces challenges including the computational cost of cryptographic verification, the need for standardized data formats across institutions, and the complexity of managing consent and access rights through smart contracts. These technical and operational considerations must be addressed before widespread adoption in clinical settings.

3.4 Financial Services and Accounting

In financial services, blockchain-AI integration addresses challenges in transaction verification, fraud detection, and regulatory compliance. Blockchain provides an immutable transaction ledger, while AI contributes pattern recognition and anomaly detection capabilities. This combination has been applied to streamline Know Your Customer (KYC) and Anti-Money Laundering (AML) processes, where AI algorithms analyze transaction patterns recorded on blockchain to identify suspicious activities [10]. For financial accounting, the integration enables automation of repetitive tasks such as transaction reconciliation and audit trail verification, potentially reducing manual processing costs and enabling more frequent financial reporting [11]. However, the practical adoption of these technologies in regulated financial environments requires addressing concerns about algorithmic accountability, the legal status of smart contract-executed transactions, and the integration of blockchain systems with existing financial infrastructure.

4 TECHNOLOGICAL AND GOVERNANCE CHALLENGES

4.1 Scalability and Performance

Scalability and performance limitations represent fundamental challenges for blockchain-AI integration. Current blockchain implementations exhibit transaction throughput and latency characteristics that may not align with the requirements of many AI applications. While optimized consensus mechanisms such as Proof of Authority can achieve approximately 1000 transactions per second in favorable conditions [3], many AI applications—particularly those involving continuous data streaming, high-frequency model updates, or real-time inference—may require substantially higher throughput. For latency-sensitive applications, blockchain transaction confirmation times present additional constraints. Although latency below 200 milliseconds has been demonstrated in specific cybersecurity implementations [4], applications requiring sub-100-millisecond responses may find blockchain verification overhead prohibitive.

Several technical approaches aim to address these limitations. Layer-2 scaling solutions, which process transactions off-chain while anchoring periodic state commitments to the main blockchain, can potentially increase effective throughput. Specialized consensus mechanisms optimized for AI workloads, rather than general-purpose transaction processing, represent another research direction. Additionally, hybrid architectures that selectively use blockchain for high-value operations (such as model version control or governance decisions) while conducting compute-intensive operations off-chain may offer practical compromises between decentralization benefits and performance requirements.

4.2 Privacy and Security

Privacy and security considerations present complex trade-offs in blockchain-AI integration. While blockchain provides tamper-resistant audit trails and cryptographic access control, the transparency inherent in public blockchains can conflict with privacy requirements for sensitive training data. Conversely, private or permissioned blockchains sacrifice some decentralization benefits to maintain confidentiality.

Privacy-preserving techniques for AI training on blockchain platforms include differential privacy, which adds calibrated noise to training data or model updates to prevent identification of individual records, and federated learning, which enables model training on distributed data without centralizing raw datasets. Research has demonstrated the integration of differential privacy with blockchain-based federated learning [12], though this approach involves inherent trade-offs: stronger privacy protections typically reduce model accuracy, requiring careful calibration based on application requirements. Emerging cryptographic techniques, including zero-knowledge proofs that enable verification of computation correctness without revealing inputs, represent additional privacy-preserving directions under active investigation.

Storage architecture also impacts security. Storing large training datasets directly on blockchain is economically impractical due to high replication costs across network nodes. Hybrid approaches that store data off-chain while recording cryptographic hashes or metadata on-chain provide verification capabilities at reduced cost [3], though they

introduce dependencies on off-chain storage systems and require careful management of access control at the boundary between on-chain and off-chain components.

4.3 Decentralized Governance Frameworks

Implementing decentralized governance frameworks for collaborative AI development presents multifaceted challenges spanning technical, economic, and social dimensions. Governance mechanisms must address several key functions: defining decision rights for model updates and policy changes, designing incentive structures that encourage quality contributions while discouraging malicious behavior, and establishing accountability mechanisms for addressing harms from model outputs.

Incentive design represents a critical challenge. Token-based reward systems can compensate participants for providing training data, computational resources, or model improvements, but measuring the value of contributions objectively remains difficult. Data quality cannot be fully verified through automated means, creating opportunities for participants to submit low-quality or adversarial data if rewards are based solely on quantity. Model performance improvements depend on complex interactions between dataset characteristics, training procedures, and evaluation metrics, complicating the attribution of credit to individual contributors. While various projects have explored different incentive structures [2,7], no consensus has emerged on optimal approaches, and designs likely need customization for specific use cases.

Transparency and interpretability present additional governance challenges. Blockchain provides transparency of transaction histories and governance votes, but this differs from the interpretability of AI decision-making processes. Recording complete model training procedures on-chain is generally impractical due to data volume, necessitating selective recording of key parameters, dataset identifiers, and validation results. Developing standards for what aspects of AI development should be recorded on-chain, and how this information should be structured for meaningful audit, remains an open research question. Furthermore, governance processes must balance inclusivity with efficiency: overly inclusive decision-making may slow development velocity, while concentrated decision-making reproduces the centralization that decentralized approaches aim to address.

4.4 Regulation and Standardization

Regulatory compliance presents significant challenges for blockchain-AI integration, particularly regarding data protection and algorithmic accountability. The General Data Protection Regulation (GDPR) in Europe establishes a "right to be forgotten," allowing individuals to request deletion of their personal data. This requirement conflicts with blockchain's immutability, which is fundamental to its security model. Potential architectural solutions include storing personal data off-chain while recording only cryptographic hashes on-chain, enabling data deletion without compromising blockchain integrity, though this approach introduces dependencies on off-chain storage systems. Alternatively, advanced cryptographic techniques such as zero-knowledge proofs or programmable privacy mechanisms could enable selective data disclosure while maintaining on-chain records. However, the legal sufficiency of these technical approaches remains uncertain, as regulatory interpretation continues to evolve. Additionally, transparency of model training and validation, while potentially beneficial for regulatory audit, must be balanced against intellectual property concerns and competitive dynamics.

Lack of standardization represents another barrier to broader adoption. Current blockchain-AI integration projects employ diverse technical approaches regarding consensus mechanisms, smart contract interfaces, data schemas, and interoperability protocols. This heterogeneity limits the ability to transfer models, datasets, or governance mechanisms across different blockchain platforms. Developing standardized protocols and APIs would facilitate interoperability and reduce implementation complexity. However, premature standardization in a rapidly evolving field risks locking in suboptimal designs. The balance between encouraging innovation through diversity and enabling interoperability through standardization remains an ongoing challenge for the research community and industry stakeholders.

5 CONCLUSION

This paper has surveyed recent research on blockchain and Large Language Model integration, examining technical approaches, application domains, and persistent challenges. Our analysis identifies three primary integration strategies: blockchain-based data repositories that provide verifiable data provenance and decentralized access control, smart contract-driven governance mechanisms that automate policy enforcement and incentive distribution, and decentralized marketplaces that facilitate peer-to-peer exchange of AI resources. Applications in cybersecurity, healthcare, and financial services demonstrate benefits particularly where data sovereignty, multi-stakeholder governance, or audit trail requirements are critical.

However, technical and governance challenges constrain practical adoption. Scalability limitations necessitate hybrid architectures balancing on-chain verification with off-chain computation. Privacy-preserving techniques involve trade-offs between data protection and model utility. Decentralized governance mechanisms face challenges in measuring contribution quality and balancing participation with efficiency. Regulatory frameworks have yet to reconcile blockchain immutability with data protection mandates such as GDPR's right to erasure.

Future research should prioritize Layer-2 scaling solutions and consensus mechanisms optimized for AI workloads, advance cryptographic techniques such as zero-knowledge proofs for privacy-preserving model training, and conduct

empirical evaluations of governance mechanisms across diverse use cases. Interdisciplinary collaboration among computer scientists, economists, legal scholars, and ethicists remains essential to address technical feasibility, economic sustainability, and regulatory compliance. This survey's limitations include its focus on English-language publications and the nascent stage of production deployments, which limits availability of real-world performance data.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This article is the teaching reform research project of Hainan Vocational University of Science and Technology in 2023: Exploration and Practice of the Application of Blockchain Technology in the Cultivation of Composite Talents in the Context of the Free Trade Port (Project No. HKJG2023-18).

REFERENCES

[1] Liu Y, Lu Q, Zhu L, et al. Decentralised Governance-Driven Architecture for Designing Foundation Model based Systems: Exploring the Role of Blockchain in Responsible AI. arXiv preprint arXiv: 2308.05962, 2023.
[2] Wang Z, Sun R, Lui E, et al. AIArena: A blockchain-based decentralized AI training platform. Companion Proceedings of the ACM on Web Conference 2025, New York, NY, USA. 2025: 1375-1379. DOI: https://doi.org/10.1145/3701716.3715484.
[3] Balachandar S K, Prema K, Kamarajapandian P, et al. Blockchain-enabled Data Governance Framework for Enhancing Security and Efficiency in Multi-Cloud Environments through Ethereum, IPFS, and Cloud Infrastructure Integration. Journal of Electrical Systems, 2024, 20(5): 2132-2139.
[4] Goundar S. Blockchain-AI Integration for Resilient Real-time Cyber Security. Global Congress on Emerging Technologies (GCET-2024), Gran Canaria, Spain. 2024: 342-349. DOI: 10.1109/GCET64327.2024.10934609.
[5] Das A K, Tonoy M T A, Hossain M. Blockchain-Based Knowledge Repository for Training Artificial Intelligence Models: Bridging AIML with Decentralized Data. 2024 IEEE Region 10 Symposium (TENSYMP), New Delhi, India. 2024: 1-6. DOI: 10.1109/TENSYMP61132.2024.10752113.
[6] Dave M, Saraf A, Kumar R, et al. DataHarbour: Enabling Decentalized AI Data Marketplace using Blockchain. 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Pune, India. 2024: 1-4. DOI: 10.1109/ICBDS61829.2024.10837338.
[7] Pisano M T, Patterson C J, Seneviratne O. PredictChain: Empowering Collaboration and Data Accessibility for AI in a Decentralized Blockchain-based Marketplace. arXiv preprint arXiv:2307.15168, 2023.
[8] Jumagaliyeva A, Abdykerimova E, Turkmenbayev A, et al. Identifying Patterns and Mechanisms of AI Integration in Blockchain for E-voting Network Security. Eastern-European Journal of Enterprise Technologies, 2024, 130(2).
[9] Alsamhi S H, Myrzashova R, Hawbani A, et al. Federated Learning Meets Blockchain in Decentralized Data Sharing: Healthcare Use Case. IEEE Internet of Things Journal, 2024, 11(11): 19602-19615.
[10] Rane N, Choudhary S, Rane J. Blockchain and Artificial Intelligence (AI) integration for revolutionizing security and transparency in finance. Available at SSRN 4644253, 2023.
[11] Kanaparthi V. Exploring the impact of blockchain, AI, and ML on financial accounting efficiency and transformation. International Conference on Multi-Strategy Learning Environment, 2024: 353-370.
[12] Xu M, Zou Z, Cheng Y, et al. SPDL: A blockchain-enabled secure and privacy-preserving decentralized learning system. IEEE Transactions on Computers, 2022, 72(2): 548-558.