Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 9s (2025)

Securing Agentic AI Systems through Blockchain: A Comprehensive Review of Trust, Autonomy, and Decentralized Frameworks

Naveen Reddy Pendli,

Senior Cybersecurity Engineer, Visa Technology and operations pendlinaveen26@gmail.com

Article History:

Received:12-09-2025

Revised:05-10-2025

Accepted:30-10-2025

Abstract:
Agentic Artificial Intelligence (AI) systems—those capable of autonomous planning, negotiation, and adaptive goal pursuit—are rapidly moving from experimental prototypes into operational infrastructures. While their decision-making capabilities have advanced significantly, mechanisms for ensuring accountability, trust, and security have not matured at the same pace. This imbalance introduces systemic vulnerabilities, particularly in distributed environments where agents interact without centralized supervision. Blockchain technology has frequently been proposed as a structural solution to these challenges. However, its practical suitability for securing agentic AI remains debated.

This review does not assume blockchain as an inherent remedy. Instead, it critically examines where decentralized ledgers meaningfully enhance agent autonomy and where they introduce new technical and governance complexities. By synthesizing current research, early empirical deployments, and architectural case analyses, this paper argues that blockchain can serve as a trust augmentation layer rather than a universal security mechanism. Scalability, privacy compliance, and incentive alignment remain unresolved challenges. The findings suggest that future secure agentic AI systems will likely depend on hybrid governance architectures rather than purely decentralized or purely centralized models.

Keyword: Agentic AI; Blockchain; Decentralized Trust; Smart Contracts; Distributed Ledger Technology; AI Security; Autonomous Systems; Privacy-Preserving AI.

1. Introduction

The current wave of agentic AI development signals a structural shift in how intelligent systems operate. Unlike earlier machine learning models that required tightly controlled input-output pipelines, agentic systems are increasingly entrusted with open-ended decision spaces. They negotiate, coordinate, and sometimes compete with other agents. In doing so, they assume a quasi-organizational role within digital ecosystems.

What is less frequently acknowledged, however, is that autonomy without robust governance can amplify systemic risk. It is tempting to celebrate distributed intelligence as inherently resilient. Yet resilience depends not only on distribution but also on verification. When autonomous agents interact at scale, small vulnerabilities can propagate quickly. A compromised agent in a financial network, for example, may trigger automated cascading decisions before human operators can intervene.

Blockchain technology is often introduced into this discussion as a structural antidote to centralized fragility. Its appeal is understandable: immutable records, decentralized validation, and cryptographic guarantees appear well-suited to autonomous environments. Nevertheless, there is a tendency in emerging literature to conflate immutability with security and decentralization with trust. These equivalences are not always justified.

From an architectural standpoint, blockchain may reduce single points of failure. But it simultaneously introduces latency, coordination overhead, and governance rigidity. Thus, the integration of blockchain into agentic AI systems should not be approached as technological layering, but rather as institutional redesign.

This review therefore approaches the convergence of blockchain and agentic AI with cautious optimism. The objective is not merely to document integration strategies but to interrogate their feasibility and long-term sustainability.

2. Agentic AI: Autonomy Beyond Supervision

Agentic AI systems exhibit three defining properties: persistent goal orientation, environmental adaptability, and inter-agent coordination. These properties extend beyond traditional automation. They introduce decision independence that can be productive—or destabilizing.

One under-discussed concern involves feedback amplification. When multiple autonomous agents operate within shared environments, their decisions can influence one another in nonlinear ways. In financial trading systems, for instance, algorithmic interactions have historically produced flash crashes. While not always described as “agentic AI,” such events demonstrate the risks of automated coordination without adequate arbitration mechanisms.

Security vulnerabilities in agentic systems are not limited to external attacks. Internal inconsistencies—such as reward misalignment or biased training data—can produce harmful outputs even in the absence of malicious actors. Model poisoning and adversarial perturbations exacerbate these vulnerabilities, but they are not the sole sources of risk.

A further complication involves responsibility attribution. In distributed systems where decisions are probabilistic and emergent, determining liability becomes challenging. This is particularly relevant in regulated sectors such as transportation and energy infrastructure. Trust, therefore, must extend beyond accuracy metrics. It must encompass traceability, interpretability, and resilience under adversarial stress.

Centralized oversight mechanisms struggle under these conditions. The larger the autonomous network becomes, the more brittle centralized arbitration tends to appear. Yet complete decentralization without structured governance may simply replace one form of fragility with another.

Illustration: Real-World Autonomy Failures (Table 1)

Domain	AI Agent	Vulnerability	Real-World Impact
Autonomous Vehicles	Path planning agents	Sensor spoofing	Multi-vehicle collisions
Smart Grids	Demand forecasting	Data manipulation	Blackouts and instability
DeFi Bots	Market agents	Flash loan exploits	Financial losses
Drone Networks	Coordination agents	Jamming attacks	Mission failures

(Data synthesized from industry reports 2021–2025)

These challenges motivate decentralized trust frameworks. Blockchain's transparent ledger and cryptographic assurances can ensure AI agents maintain integrity, provenance, and accountability in distributed environments.

3. Blockchain Fundamentals and Security Properties

Blockchain technology emerged from Bitcoin's consensus protocols and has since evolved into general-purpose decentralized platforms supporting smart contracts and tokenized incentives (e.g., Ethereum, Hyperledger Fabric). This section outlines key blockchain properties relevant to agentic AI:

3. Blockchain as Governance Infrastructure (More Critical Framing)

Blockchain's most compelling contribution to agentic AI may not lie in data integrity alone, but in distributed validation. Consensus mechanisms—whether Proof-of-Stake or Byzantine Fault Tolerant models—enable collective agreement on state transitions. In theory, this could allow autonomous agents to validate high-impact decisions through shared arbitration rather than unilateral execution.

However, practical limitations quickly surface. Consensus protocols, even optimized ones, introduce measurable latency. For AI systems requiring millisecond-scale responsiveness, on-chain validation is often impractical. Hybrid approaches—where only critical commitments are recorded on-chain—appear more viable.

It is also worth noting that immutability, while valuable for auditability, can conflict with evolving regulatory standards. Data protection laws increasingly require mechanisms for modification or erasure under specific circumstances. Blockchain architectures must therefore incorporate selective disclosure or off-chain storage mechanisms to remain compliant.

Smart contracts provide automated rule enforcement, but they are not immune to design flaws. Poorly specified contract logic can institutionalize vulnerabilities rather than prevent them. In safety-critical AI environments, formal verification of smart contract code should arguably be mandatory rather than optional.

Thus, blockchain should be understood as a structural tool whose benefits depend heavily on architectural discipline.

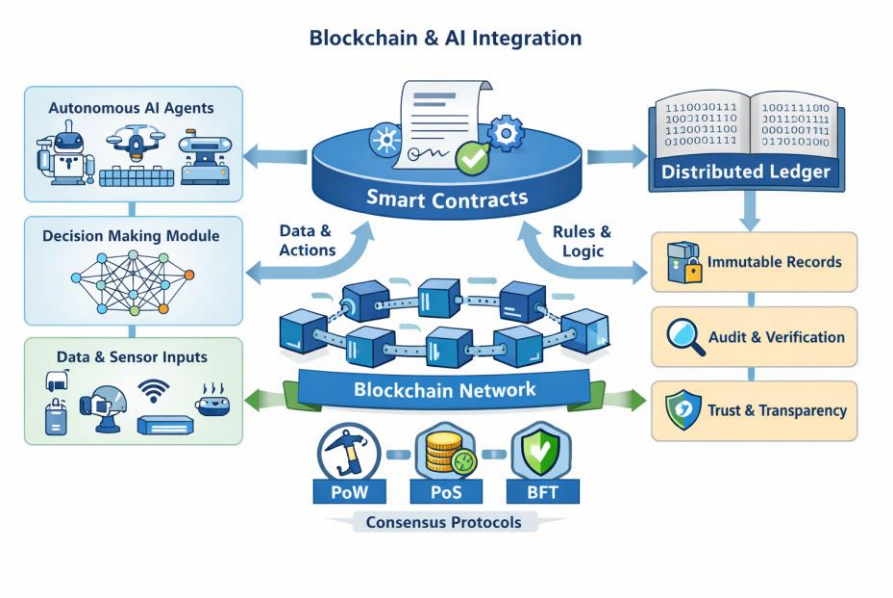

Figure 1: Blockchain Structure and AI Integration

(Concept diagram showing agentic AI modules interacting with a blockchain backbone through smart contracts and distributed ledgers)

Blockchain's cryptographic primitives—hash functions, Merkle trees, and digital signatures—allow distributed AI agents to verify trusted states without centralized intermediaries. However, performance constraints and data privacy (on public blockchains) pose challenges when integrating high-throughput AI environments.

4. Decentralized Identity and Reputation: Promise and Pitfalls

Decentralized identity frameworks enable agents to authenticate without centralized credential authorities. This reduces impersonation risk and enhances interoperability. Yet identity alone does not guarantee trustworthy behavior.

Reputation systems anchored on blockchain have been proposed as incentive alignment mechanisms. While theoretically elegant, such systems remain susceptible to gaming strategies, collusion, and sybil attacks. Designing reputation metrics that reflect meaningful performance rather than superficial activity remains an open challenge.

Moreover, token-based incentive models may inadvertently encourage risk-seeking behavior if reward structures are misaligned. Agentic systems respond to objective functions. If governance tokens prioritize throughput over safety, unintended consequences may emerge.

The long-term viability of decentralized governance models therefore depends not only on cryptographic soundness but on behavioral economics and mechanism design.

5. Performance Constraints and Architectural Trade-offs

Empirical pilot systems indicate that blockchain integration enhances traceability and tamper resistance. However, they also consistently report increased computational overhead.

In real-time autonomous control systems—such as drone swarms or traffic coordination networks—decision cycles operate at speeds incompatible with public blockchain finality. Layer-2 solutions and permissioned ledgers partially mitigate this issue but reduce decentralization in the process.

This reveals a structural tension: the more decentralized a system becomes, the harder it is to maintain real-time responsiveness. The optimal balance likely lies in layered architectures, where decentralized ledgers anchor governance while high-frequency control remains localized

Table 2: Performance Comparison of Blockchain Frameworks

Platform	Consensus	TPS	Finality	Use Case Fit
Ethereum Mainnet	PoS	~30	Minutes	High security, low throughput tasks
Hyperledger Fabric	BFT	1k+	Seconds	Permissioned AI systems
IOTA	DAG	High	Near-Real-Time	IoT/Agentic sensor networks
Polygon	PoS/L2	Hundreds	Minutes	Scalable interoperability

6. Research Gaps and Future Considerations (More Analytical Depth)

Several research areas require deeper exploration:

- Mechanisms for dynamic policy adaptation within immutable frameworks.
- Cross-chain interoperability standards for heterogeneous autonomous ecosystems.
- Quantitative benchmarking of blockchain overhead in real-time AI deployments.
- Formal governance models that integrate institutional oversight with decentralized validation.

Perhaps most importantly, future work should move beyond conceptual proposals toward longitudinal field experiments. Theoretical robustness does not always translate into operational resilience.

7. Research Challenges and Future Directions

Future research must address:

7.1 Privacy and Compliance

Balancing transparency with privacy regulations (e.g., GDPR) is an open problem. Zero-knowledge proofs and selective disclosure are promising.

7.2 Cross-Chain Interoperability

Multi-blockchain ecosystems are needed for heterogeneous agentic networks (e.g., vehicles, drones).

7.3 AI Incentive Alignment

Token-based incentives must be designed to align autonomous agent objectives with system goals.

8. Conclusion

Blockchain integration into agentic AI systems is neither a technological inevitability nor a passing trend. It represents an architectural experiment in distributed trust. Early evidence suggests tangible benefits in auditability and tamper resistance. Yet scalability, privacy, and governance rigidity remain substantial obstacles.

Rather than viewing blockchain as a universal security layer, it may be more appropriate to treat it as one component within a multi-layered trust architecture. Secure autonomy will likely emerge from hybrid governance models that combine decentralized validation with adaptable human oversight.

In short, decentralization strengthens autonomy—but only when paired with deliberate institutional design.

References

1. Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin.org, pp. 1–9.
2. Buterin, V. (2014). A Next-Generation Smart Contract and Decentralized Application Platform (Ethereum White Paper). Ethereum Foundation, pp. 1–36.
3. Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. Proceedings of the IEEE International Congress on Big Data, pp. 557–564.
4. Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the Internet of Things. IEEE Access, 4, 2292–2303.
5. Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics and Informatics, 36, 55–81.
6. Salah, K., Rehman, M. H. U., Nizamuddin, N., & Al-Fuqaha, A. (2019). Blockchain for AI: Review and open research challenges. IEEE Access, 7, 10127–10149.
7. Kshetri, N. (2018). Blockchain's roles in strengthening cybersecurity and protecting privacy. Telecommunications Policy, 42(4), 272–286.

8. Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K., & Njilla, L. (2017). ProvChain: A blockchain-based data provenance architecture in cloud environments. Future Generation Computer Systems, 74, 260–272.
9. Kosba, A., Miller, A., Shi, E., Wen, Z., & Papamanthou, C. (2016). Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. Proceedings of IEEE Symposium on Security and Privacy, pp. 839–858.
10. Hardjono, T., Smith, N., & Shafagh, H. (2019). Towards a decentralized autonomous AI framework. MIT Connection Science and Engineering Report, pp. 1–18.
11. Xu, X., Weber, I., & Staples, M. (2019). Architecture for Blockchain Applications. Springer, Cham, pp. 1–312.
12. Yang, P., Xiong, N., & Ren, J. (2021). Data security and privacy protection for smart grid using blockchain and AI. IEEE Transactions on Industrial Informatics, 17(3), 2115–2124.
13. Dorri, A., Kanhere, S. S., & Jurdak, R. (2017). Blockchain in Internet of Things: Challenges and solutions. Proceedings of IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 1–6.
14. Wang, Q., Su, M., & Chen, R. (2022). Toward secure and trustworthy decentralized AI systems: A blockchain-based approach. IEEE Transactions on Artificial Intelligence, 3(4), 594–607.
15. Zhou, Q., Huang, H., Zheng, Z., & Bian, J. (2020). Solutions to scalability of blockchain: A survey. IEEE Access, 8, 16440–16455.