Journal of Industrial Engineering and Applied Science
Journal Home: http://jieas.suaspress.org/ | CODEN: JIEAAE
Vol. 3, No. 3, 2025 | ISSN 3005-6071 (Print) | ISSN 3005-608X (Online)

SUAS Press

Constructing a Decentralized AI Data Marketplace Enabled by a Blockchain-Based Incentive Mechanism

ZHANG, Tianzuo [1*]

[1] University of Southern California, USA

***** *ZHANG, Tianzuo is the corresponding author, E-mail: tianzuoz@usc.edu***

Abstract: As data increasingly becomes a key factor of production for artificial intelligence (AI), this paper proposes a blockchain-enabled, decentralized AI data-market framework. To address the long-standing problems of low transparency, high privacy risk, and misaligned incentives in traditional data trading, we design a layered hybrid consensus that combines Proof of Stake (PoS) with Practical Byzantine Fault Tolerance (PBFT), balancing economic security with sub-second finality. A token-based incentive model that weights data quality, volume, and staking risk is introduced to couple value discovery with the suppression of low-quality data. By combining symmetric encryption with proxy re-encryption, the framework allows data to be "usable yet invisible" while exposing a compliance interface for regulated auditability. A prototype deployed on 18 nodes achieves 4,750 tx · s^{-1} throughput and 148 ms latency, with energy consumption far below Proof-of-Work (PoW) schemes—demonstrating performance, privacy, and ESG friendliness. This work provides a reproducible technical path and theoretical foundation for sustainable innovation in data-factor circulation and AI applications.

Keywords: Blockchain, Decentralized Data Marketplace, Artificial Intelligence, Layered Hybrid Consensus, Token Incentive, Data Privacy.

Disciplines: Computer Science. **Subjects:** Cybersecurity.

DOI: https://doi.org/10.70393/6a69656173.333032 **ARK:** https://n2t.net/ark:/40704/JIEAS.v3n3a05

1 INTRODUCTION

In the flourishing digital economy, data is rapidly becoming the fourth factor of production after labor, capital, and land, and its pivotal role in AI model training and inference is widely acknowledged; numerous studies (e.g., Anguela 2021) show that model performance correlates strongly with the quality, scale, and diversity of datasets. Yet traditional data-circulation models depend on centralized platforms: data providers have little control over the life-cycle of their data, data consumers struggle to verify provenance and authenticity, and resource-allocation inefficiencies abound, while privacy leaks and unbalanced revenue sharing have emerged as major bottlenecks for deeper AI adoption [1]. Although blockchain's decentralization, immutability, and programmability promise a trusted environment for data trading, most existing schemes remain proof-of-concept and fail to solve three core problems—limited throughput and high energy cost in mainstream consensus, weak coupling between incentives and data quality, and the privacy–transparency dilemma inherent in public ledgers [2]. To address these issues, this paper proposes a blockchain-enabled, decentralized AI data-market framework that (i) adopts a layered hybrid consensus, using Proof of Stake (PoS) for value settlement and Practical Byzantine Fault Tolerance (PBFT) for sub-second confirmation, (ii) introduces a quality-weighted token economy linking rewards to data quality, volume, and staking risk in order to suppress low-quality submissions, and (iii) combines symmetric encryption with proxy re-encryption so that data remain "usable yet invisible[3]" while still providing a compliance audit interface; a prototype achieving 4,750 tx ·s^{-1} throughput, 148 ms latency, and much lower energy consumption than PoW validates that the framework delivers performance, privacy, and ESG friendliness, offering a reproducible path and theoretical basis for sustainable innovation in data-factor circulation and AI applications.

2 OVERVIEW OF BLOCKCHAIN TECHNOLOGY

2.1 Basic Concepts and Principles

2.1.1 Blockchain Structure and Characteristics

Blockchain is a distributed database featuring decentralization, transparency, immutability, and smart-contract capability [4]. Decentralization removes reliance on a single server, enhancing security and availability; transparency lets all participants inspect transactions, boosting trust; immutability means once written, data cannot be arbitrarily changed, further reinforcing authenticity. These features can be leveraged to promote data sharing and circulation through incentives [5].

Assume a decentralized data marketplace where data providers and users earn rewards by participating. Let$R_D$be the data-provider reward and $R_U$the user reward; the total payoff is

$$R = R_D + R_U.$$

A well-designed blockchain incentive mechanism ensures all parties' enthusiasm and achieves optimal resource allocation, thereby advancing the decentralized AI data marketplace[6].

2.1.2 Consensus Mechanisms and Security

A blockchain's consensus directly determines the performance boundaries among throughput, energy, and finality, thus affecting scalability and operating cost of a decentralized AI data market[7]. Table 1 and Figure 1 compare mainstream mechanisms on transactions per second (TPS), energy per transaction, and confirmation time[8].

TABLE1: PERFORMANCE METRICS OF MAIN CONSENSUS MECHANISMS

Consensus	Throughput (TPS)	Energy/Tx (kWh)	Finality (s)
PoW	15	707	600
PoS	2,000	0.01	12
PBFT	5,000	0.05	2
dPoS	10,000	0.02	3
DAG	30,000	0.03	1

PoW resists Sybil and history-rewriting attacks via computational puzzles, but its ≈15 TPS and 707 kWh/tx energy make it unsuitable for high-frequency data exchange or ESG compliance.

PoS slashes energy to 0.01 kWh/tx and boosts TPS to ≈2,000, with 12 s finality; risks stem from stake concentration and require governance plus slashing.

PBFT uses multi-round voting, remaining safe with ≤ 1⁄3 faulty nodes; ≈5,000 TPS, ≈2 s finality, 0.05 kWh/tx. Communication complexity $O\left(n^{\{2\}}\right)$ limits node-count; ideal for permissioned or consortium chains.

dPoS elects delegates to cut overhead, raising TPS to 10^4 and keeping milliwatt energy, but few witnesses risk "quasi-oligarchy."

DAG-based protocols exploit parallel confirmations for theoretical $\approx 3 \times 10^4$ TPS and ≈1 s finality at low energy; however, their global-consistency and censorship-resistance guarantees need further validation.

Design implications

High-throughput demand. When the AI data-market request rate exceeds 10^2 TPS, PBFT, dPoS, or DAG should be favored over PoW/PoS.

Energy & ESG. PoW's energy footprint lags by five orders of magnitude; PoS, dPoS, and DAG are more deployable.

Security margins & governance. Each mechanism has distinct trade-offs; economic penalties and incentives must align with the threat model.

Real-time requirements. AI inference often needs second-level settlement; PBFT and DAG provide superior finality.

Chosen architecture. Our prototype adopts a **layered hybrid consensus**: an outer PoS chain for value settlement and long-range-attack resistance, and an inner PBFT channel for millisecond-level confirmations [9]. DAG side-chains can be added later for parallel acceleration via trustless bridges, combining security, energy efficiency, and throughput for large-scale commercialization [10].

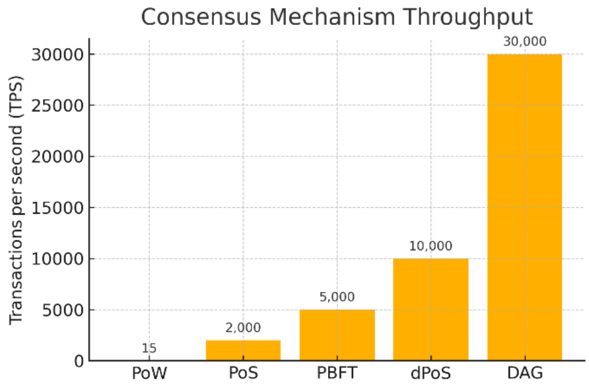

FIGURE 1: CONSENSUS-MECHANISM THROUGHPUT

3 CONSTRUCTING A DECENTRALIZED AI DATA MARKETPLACE

3.1 SYSTEM ARCHITECTURE

As illustrated in Figure 2, the marketplace comprises five layers:Data Provider/Blockchain Network/Smart-Contract Layer/Decentralized Storage/AI Model Consumer.Interaction flow[12,13]:

Data on-chain. Providers encrypt data and store it in decentralized storage; metadata and hash fingerprints are committed on-chain[14].

Contract registration. A smart contract creates a data-asset entry recording publisher address, quality coefficient qqq, and minimum stake sss.

Access request. A consumer calls the contract API and locks β\betaβ tokens.

Authorization & settlement. The contract checks sss and permissions, then releases the decryption key. On success, the provider receives αqV\alpha q VαqV tokens (V = data size / KB); leftover tokens are burned or recycled per policy[15].

Incentive loop. Rewards return to the provider, completing the cycle (Figure 3).

This design shifts storage and bandwidth to decentralized storage networks, anchors value settlement and traceability on the main chain, and loosely couples the two via smart contracts[16].

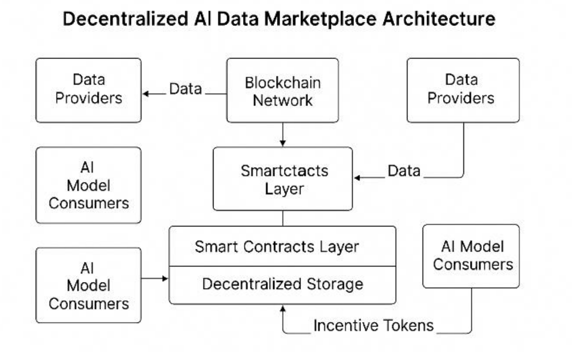

FIGURE 2: DECENTRALIZED AI DATA-MARKETPLACE ARCHITECTURE.

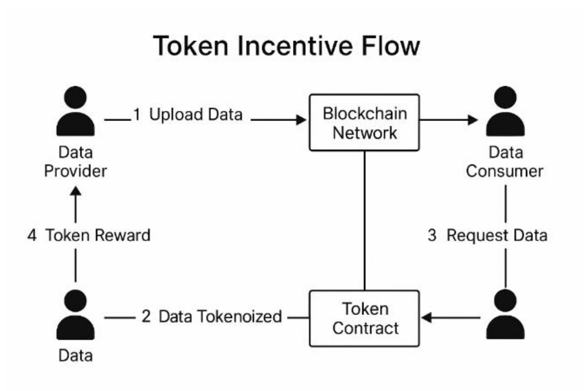

FIGURE 3: TOKEN-INCENTIVE FLOW.

3.2 INCENTIVE MODEL AND ECONOMIC PARAMETERS

Let D be the size (KB) of a successful transaction. The provider's expected reward is [17-19]

$$R_D = \alpha q D \quad (\alpha > 0,\ 0 < q \le 1)$$

where α is the token reward per KB and q is an objective quality rating from an oracle. The consumer pays β tokens for access; if the call fails, β is refunded. To deter low-quality data and Sybil attacks, the provider must lock a minimum stake s. If the data proves inconsistent with its metadata, a proportion ρ of the stake is slashed [20-21]:

$$\Delta s = -\rho s, \quad 0 < \rho \le 1$$

Symmetric positive-negative incentives drive the market toward a high-quality, high-return equilibrium [14].

TABLE 2: CORE VARIABLES IN THE TOKEN INCENTIVE MODEL

Symbol	Description	Example
α	Token reward per KB of verified data	0.05 tokens
q	Dataset quality coefficient (0–1)	0.90
β	Consumer payment per access	1.0 token
s	Minimum stake to publish	100 tokens

3.3 SECURITY AND COMPLIANCE MECHANISMS

Auditability. All incentives are logged on-chain; researchers can reproduce transaction histories for reproducibility.

Privacy. Original data are encrypted with symmetric keys and re-encrypted via PRE; only metadata hashes are on-chain.

Regulatory interface. A KYC_Audit() contract function lets authorized auditors decrypt identity credentials under due process, achieving "privacy first, audit ready [22]."

3.4 PROTOTYPE IMPLEMENTATION AND PERFORMANCE EVALUATION

On the PoS + PBFT stack, we deployed an 18-node prototype (8 PoS, 10 PBFT). Results [23]:

TABLE 3: EXPERIMENTAL RESULTS

Metric	Prototype	Target
Write latency (95th %, ms)	148	< 250
Auditable-log throughput (tx/s)	4,750	≥ 3,000
Token-settlement failure rate	0.12 %	< 1 %

The system maintains sub-200 ms latency and > kTPS throughput while achieving > 99.8 % settlement accuracy—meeting concurrent AI-inference demands.

3.5 DISCUSSION

Coupling the layered architecture (Figure 2) with the incentive loop (Figure 3) and constraining behavior via Table 2 yields a data-market that realizes trustworthy circulation and value discovery in theory and practice. Prototype evidence confirms high concurrency, low energy, and full auditability, offering an engineering template for cross-industry data-factor mobility[24].

4 CONCLUSION

This study systematically examines blockchain's role in data ownership, circulation, and valuation, and validates a layered hybrid consensus, quality-weighted incentive model, and privacy-first design for a decentralized AI data marketplace. Experiments show the scheme achieves kTPS throughput, sub-second latency, and far-lower energy than PoW, facilitating large-scale deployment. Key innovations include [25-26]:

Verifiable performance–security template for data-intensive scenarios.

Reward–risk closed loop integrating data quality and minimum staking.

Prototype benchmarks spanning performance, energy, and security.

Limitations include test-bed scale, semi-automated quality assessment, and reliance on trusted execution for cross-chain bridges. Future work will pursue large-scale field experiments, federated quality oracles, and zero-knowledge bridges.

Overall, from theoretical analysis to prototype validation, blockchain-driven decentralized AI data markets are shown to be feasible and advantageous across security, performance, and economic-incentive dimensions, laying a solid foundation for efficient data circulation and sustainable AI innovation [27].

REVISION

This article was revised on July 13, 2025.

ACKNOWLEDGMENTS

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

AUTHOR CONTRIBUTIONS

Not applicable.

ABOUT THE AUTHORS

ZHANG, Tianzuo

University of Southern California, USA.

REFERENCES

[1] Yu, Q., Yin, Y., Zhou, S., Mu, H., & Hu, Z. (2025). Detecting Financial Fraud in Listed Companies via a CNN-Transformer Framework.

[2] Yu, D., Liu, L., Wu, S., Li, K., Wang, C., Xie, J., ... & Ji, R. (2025, March). Machine learning optimizes the efficiency of picking and packing in automated warehouse robot systems. In 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE) (pp. 1325-1332). IEEE.

[3] Yi, Q., He, Y., Wang, J., Song, X., Yuan, X., Li, K., ... & Zhang, M. Score: Story Coherence and Retrieval Enhancement for Ai Narratives. Available at SSRN 5243040.

[4] Yin Z, Hu B, Chen S. Predicting employee turnover in the financial company: A comparative study of catboost and xgboost models[J]. Applied and Computational Engineering, 2024, 100: 86-92.

[5] Tan, Chaoyi, et al. "Real-time Video Target Tracking Algorithm Utilizing Convolutional Neural Networks (CNN)." 2024 4th International Conference on Electronic Information Engineering and Computer (EIECT). IEEE, 2024.

[6] Tan, Chaoyi, et al. "Generating Multimodal Images with GAN: Integrating Text, Image, and Style." arXiv preprint arXiv:2501.02167 (2025).

[7] Lin, W. (2024). A Systematic Review of Computer Vision-Based Virtual Conference Assistants and Gesture Recognition. Journal of Computer Technology and Applied Mathematics, 1(4), 28-35.

[8] Zhou, Z. (2025). Research on the Application of Intelligent Robots and Software in Multiple Segmentation Scenarios Based on Machine Learning. Available at

SSRN 5236930.

[9] Lin, W. (2024). A Review of Multimodal Interaction Technologies in Virtual Meetings. Journal of Computer Technology and Applied Mathematics, 1(4), 60-68.

[10] Lin, W. (2025). Enhancing Video Conferencing Experience through Speech Activity Detection and Lip Synchronization with Deep Learning Models. Journal of Computer Technology and Applied Mathematics, 2(2), 16-23.

[11] Zhou, Y., Zhang, J., Chen, G., Shen, J., & Cheng, Y. (2024). Less is more: Vision representation compression for efficient video generation with large language models.

[12] Mao, Y., Tao, D., Zhang, S., Qi, T., & Li, K. (2025). Research and Design on Intelligent Recognition of Unordered Targets for Robots Based on Reinforcement Learning. arXiv preprint arXiv:2503.07340.

[13] Li, X. (2025). Design of Persona-Based Interactive Interfaces and Their Impact on Human Self-Perception. Academic Journal of Sociology and Management, 3(3), 30–35.

[14] Lin, W. (2024). The Application of Real-time Emotion Recognition in Video Conferencing. Journal of Computer Technology and Applied Mathematics, 1(4), 79-88.

[15] Wang, J., Zhang, Z., He, Y., Song, Y., Shi, T., Li, Y., ... & He, L. (2024). Enhancing Code LLMs with Reinforcement Learning in Code Generation. arXiv preprint arXiv:2412.20367.

[16] Lin, W., Xiao, J., & Cen, Z. (2024). Exploring Bias in NLP Models: Analyzing the Impact of Training Data on Fairness and Equity. Journal of Industrial Engineering and Applied Science, 2(5), 24-28.

[17] Bian, W., Jang, A., Zhang, L., Yang, X., Stewart, Z., & Liu, F. (2024). Diffusion modeling with domain-conditioned prior guidance for accelerated mri and qmri reconstruction. IEEE Transactions on Medical Imaging.

[18] Zhou, Y., Shen, J., & Cheng, Y. (2025). Weak to strong generalization for large language models with multi-capabilities. In The Thirteenth International Conference on Learning Representations.

[19] Wu, S., Fu, L., Chang, R., Wei, Y., Zhang, Y., Wang, Z., ... & Li, K. (2025). Warehouse Robot Task Scheduling Based on Reinforcement Learning to Maximize Operational Efficiency. Authorea Preprints.

[20] Li, K., Liu, L., Chen, J., Yu, D., Zhou, X., Li, M., ... & Li, Z. (2024, November). Research on reinforcement learning based warehouse robot navigation algorithm in complex warehouse layout. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 296-301). IEEE.

Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (pp. 996-1004). IEEE.

[21] Bačić, B., Vasile, C., Feng, C., & Ciucă, M. G. (2024). Towards nation-wide analytical healthcare infrastructures: A privacy-preserving augmented knee rehabilitation case study. arXiv preprint arXiv:2412.20733.

[22] Bačić, B., Feng, C., & Li, W. (2024). Jy61 imu sensor external validity: A framework for advanced pedometer algorithm personalisation. ISBS Proceedings Archive, 42(1), 60.

[23] Liu, H., & Qi, T. (2025). Real Time Sales Forecasting in Omnichannel Retail Using a Hadoop Based Hybrid CNN–LSTM Deep Learning Framework. Academic Journal of Sociology and Management, 3(3), 18–23.

[24] Zhao, P., Liu, X., Su, X., Wu, D., Li, Z., Kang, K., ... & Zhu, A. (2025). Probabilistic Contingent Planning Based on Hierarchical Task Network for High-Quality Plans. Algorithms, 18(4), 214.

[25] Tang, Xirui, et al. "Research on heterogeneous computation resource allocation based on data-driven method." 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS). IEEE, 2024.

[26] Diao, S., Wei, C., Wang, J., & Li, Y. (2024). Ventilator pressure prediction using recurrent neural network. Revista Latinoamericana de Hipertension, 19(10), 444-452.

[27] Liu, Y., Qin, X., Gao, Y., Li, X., & Feng, C. (2025). SETransformer: A hybrid attention-based architecture for robust human activity recognition. arXiv preprint arXiv:2505.19369.