Babylonian Journal of Artificial Intelligence
Vol. **(2023)**, 2023, **pp**. 1–4
DOI: https://doi.org/10.58496/BJAI/2023/001 ISSN: xxxxxx
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Big Data and Artificial Intelligence on the Blockchain: A Review

Saad Abbas Abed[1],*,

[1] *Department of Computer, College of Education, Aliraqia University, Baghdad, Iraq*

Article info

Article History
Received 18 Oct 2022
Accepted 03 Jan 2023
Published 10 Jan 2023

Keywords
Artificial Intelligence
Big Data
Blockchain
Review

Abstract

As organizations seek to derive greater value from rapidly expanding datasets, they face challenges related to big data governance including security, privacy, regulatory compliance, access controls, and building trust across entities. Blockchain has emerged as a promising technology for managing and analyzing big data in a controlled, transparent manner through its decentralized architecture and cryptographic foundations. This paper explores how blockchain can enable more secure sharing of big data that mitigates systemic impediments to wider application. We first examine unique properties of blockchain technology that allow distrusting parties to collaborate through immutable records, consensus protocols, smart contracts, and permissions management. We then profile real-world use cases in sectors like healthcare, supply chain logistics, and financial services where blockchain strengthens big data usage. Finally, current scaling and architectural limitations are discussed along with an outlook on the remaining technology gaps and skills needed to fully bridge blockchain, distributed data, advanced analytics, and machine learning. The paper ultimately argues blockchain may profoundly expand big data's capabilities but still requires further innovation across hardware infrastructure, platforms, analytics interfaces, and organizational strategy.

1. INTRODUCTION

Blockchain and distributed ledger technology has emerged in recent years as a decentralized way to publicly and securely record transactions between parties without the need for a central authority [1]. Based on cryptographic principles, blockchain allows an immutable, transparent record of transactions that can be viewed by participating parties [2]. Meanwhile, the growth of big data - from sources like web traffic, social media, mobile devices, and internet of things sensors - has created tremendous opportunities for organizations to derive value through advanced analytics [3]. However, organizations also grapple with effectively storing, managing, and governing huge, rapidly increasing datasets [4]. These data governance challenges include privacy concerns, security risks, access controls, and building trust across different parties [5]. Blockchain technology has the potential to address many of these challenges and enable more controlled, transparent sharing of big datasets [6]. This paper will argue that blockchain has evolved into an important tool for not just recording transactions but also utilizing and governing big data storage and analytics across organizations [7]. By combining blockchain's decentralized approach with big data pools, we can enable more parties to share and derive value from data while maintaining critical controls like compliance, access restrictions, and participant consent [8].

2. BACKGROUND

2.1 Artificial Intelligence

Artificial intelligence (AI) and blockchain are two cutting-edge technologies that have each generated enormous interest in recent years. When combined, AI and blockchain have the potential to enable innovative new applications and transform many industries.

One of the most promising applications of AI and blockchain is decentralized AI platforms. Traditionally, AI algorithms and models have been controlled by large tech companies in centralized data centers. However, blockchain can enable fully

*Corresponding author. Email: saaa2013@yahoo.com

decentralized AI networks. These use distributed ledgers and smart contracts for coordination, data sharing, and transactions between participants. The result is collaborative training of AI models without a central authority. Decentralized AI leverages the collective data and computing resources of network participants to create models that no single party fully owns or controls. These models could have more real-world integrity as incentives are better aligned. AI can also enhance various aspects of blockchain architecture itself. This includes optimizing consensus mechanisms, improving security against cyberattacks, facilitating data analytics on blockchains, and supporting the automation of various processes. For example, machine learning algorithms could be integrated in proof-of-work networks to customize the difficulty dynamically instead of relying on fixed formulas. AI could also detect and respond to security threats by analyzing transaction patterns. Over time, blockchain systems could even leverage AI to become more self-governing.

As AI and blockchain converge, they enable new technological possibilities that leverage the strengths of both. This opens up opportunities across finance, healthcare, supply chains, government, entertainment, transportation, and more. However, challenges related to transparency, bias, and control also need to be overcome for decentralized AI to achieve its potential. Ongoing advances in the space are rapidly accelerating innovation.

2.2 Blockchain Architecture Enables Secure Data Sharing

Blockchain's decentralized architecture facilitates secure sharing of data between parties who may not fully trust each other. Rather than relying on a central intermediary, blockchain uses a distributed consensus model requiring confirmation across nodes on the network [9]. This eliminates the need for a third-party authority to oversee transactions or data exchange. Cryptographic techniques like hashing, digital signatures, and public-key infrastructure enable nodes to validate identities and transactions [10]. Once data is written to the blockchain, cryptography immutably preserves its state while still allowing permissioned access [11]. transformed. This audit trail builds integrity and trust in the data sharing process [12]. While centralized databases can be altered by their owners, blockchain's cryptographically linked blocks remain tamper-proof. These properties allow enterprises to share sensitive data like medical records, financial information, or supply chain details with other institutions, despite competitive or legal privacy concerns [13]. Access can be granted to specified transaction data without compromising security or revealing identities.These technical foundations create a transparent record showing how and when data is accessed or Overall, blockchain architecture provides a technical basis for managed data sharing across partially trusting parties requiring both privacy and collaborative analytics. The technology continues advancing through innovations like permissioned blockchains, zero-knowledge proofs, trusted execution environments, and off-chain storage solutions [14]. Together, these concepts expand blockchain's capabilities for governed data usage.

3. MANAGING BIG DATA

Advanced analytics on large, complex datasets enables organizations to uncover insights for improved decision-making. However, as volumes of data grow exponentially each year, proper governance around privacy, permissions, and information security becomes critical [15]. Though big data generates business value, uncontrolled sharing risks confidentiality and regulatory non-compliance.

Blockchain presents a robust model for standardized data governance across industries and applications [2]. Its cryptographically validated network can enforce fine-grained access policies expressed through smart contracts. These programmable agreements execute rules around data analysis, giving organizations dynamic control even after information gets shared externally [3]. For example, healthcare providers can authorize restricted third-party research queries on patient records while maintaining HIPAA compliance [16]. Additionally, blockchain empowers new data monetization models through transparent auditing of usage metrics and automated micropayments. Participants gain the ability to price data sharing according to market demand while honoring the original owner's consent rules [17]. This helps quantify the tangible value created from data analytics. Overall, blockchain architecture elevates big data governance with stricter yet more flexible controls to encourage secure multi-party collaboration.

4. REAL-WORLD USE CASES

Sensitive patient information can be analyzed to improve treatments, but sharing data risks privacy violations. Blockchain solutions enable multi-institution data queries while logging access and preserving anonymity [8]. Smart contracts automatically verify credentials and permissions prior to unlocking encrypted records. This allows aggregated population health insights across groups like cancer patients or clinical trial participants without individual consent issues.

IoT sensors now generate abundant data on time, location, temperature, and handling of food products. Blockchain ledger architecture provides transparency and accountability across global shipment pathways [17]. It cryptographically traces asset custody from farmers to processors to distributors using timestamps and GPS data. Automated alerts can trigger on environmental condition violations. This builds consumer trust and addresses liability concerns.

As financial institutions produce growing amounts of trade data, blockchain systems can securely share detailed logs with regulators for market surveillance [18]. These tamper-proof, consolidated audit trails lower compliance costs while still meeting verification mandates. All participating entities can view shared records of timestamped transactions, positions, order book data, and exchange reporting. This improves synchronization, transparency, and accountability.

5. CHALLENGES AND LIMITATIONS

1. Scalability of Permissionless Blockchains

Public blockchains like Ethereum face scalability hurdles supporting high transaction throughput and low latency needed for sizable data workloads [19]. Various solutions are under development including sharding, state channels, and directed acyclic graphs. For now, private enterprise blockchains better handle non-trivial data processing.

2. On-Chain vs Off-Chain Storage and Computing

Storing entire datasets on-chain remains expensive and inefficient compared to traditional data lakes [20]. Hybrid solutions place raw data in cloud servers while logging hashes, metadata, and transformations on-chain. But computational analysis relies on off-chain analytics engines. This split architecture has tradeoffs.

3. Oracle Problem of Linking Off-Chain Data

Connecting external data to blockchain smart contracts requires an oracle to validate real-world states. But existing oracle approaches face security vulnerabilities and single points of failure [21]. Decentralized oracle networks aim to overcome these issues through consensus mechanisms, however remain a work in progress.

4. Skills Gap

Integrating blockchain, data lakes, analytics, and machine learning requires teams skilled in both distributed ledgers and big data tools [22]. But professionals with cross-domain experience remain rare while vendor solutions stay fragmented. Substantial technology and skill investments are necessary to unify these complex platforms.

6. CONCLUSIONS

In summary, blockchain technology shows immense promise as a tool for governing big data storage, sharing, and analytics across diverse entities. By providing transparent yet access-controlled data infrastructure, blockchain architecture helps overcome systemic impediments related to privacy, security, regulatory burdens, and lack of trust that currently hinder more impactful big data applications . This paper offered several compelling use cases, from supply chains to healthcare, where blockchain systems enable secure yet compliant data usage and collaboration between partially trusting parties.

However, blockchain and big data remain rapidly evolving technologies with gaps needing ongoing research and innovation . Technical constraints around permissionless ledger throughput, on-chain storage capacities, external data integration, and cross-domain analytical skills must still be addressed by industry developers. Hybrid architectures and new cryptographic techniques show potential to mitigate these barriers over time. Ultimately, the combination of scalable, high-performance data lakes with the integrity and auditability of blockchain ledgers can profoundly expand the power of big data to drive transparency, accountability, regulatory compliance, and predictive insights across previously siloed or distrusting institutions. But fully realizing this potential will rely on patient progress across the layers of emerging hardware, blockchain platforms, distributed data schemas, analytics interfaces, and organizational best practices.

Conflicts of Interest

The author's paper clearly states that no conflicts of interest exist in relation to the research or its publication.

Funding

The author's paper explicitly states that no funding was received from any institution or sponsor.

Acknowledgment

The author acknowledges the assistance and guidance received from the institution in various aspects of this study.

References

[1] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Decentralized Business Review, 21260.
[2] Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). Blockchain technology: Beyond bitcoin. Applied Innovation, 2(6-10), 71.
[3] Einav, L., & Levin, J. (2014). Economics in the age of big data. Science, 346(6210), 1243089.

[4] Abbas, A., Bilal, K., Zhang, L., & Khan, S. U. (2015). A cloud based health insurance plan recommendation system: A big data analytics approach. Future Generation Computer Systems, 43, 99-109.

[5] Yang, Y., Lee, P., & Hajli, N. (2021). Blockchain in healthcare: A systematic literature review, synthesised framework and future research agenda. Technological Forecasting and Social Change, 162, 120412.

[6] Karafiloski, E., & Mishev, A. (2017). Blockchain solutions for big data challenges: A literature review. In IEEE EUROCON 2017-17th International Conference on Smart Technologies (pp. 763-768).

[7] Gatteschi, V., Lamberti, F., Demartini, C., Pranteda, C., & Santamaría, V. (2018). Blockchain and smart contracts for insurance: Is the technology mature enough?. Future Internet, 10(2), 20.

[8] Zhang, P., White, J., Schmidt, D. C., Lenz, G., & Rosenbloom, S. T. (2018). FHIRChain: applying blockchain to securely and scalably share clinical data. Computational and Structural Biotechnology Journal, 16, 267-278.

[9] Crosby, M. et al. (2016). Blockchain Technology: Beyond Bitcoin. Applied Innovation Review, (2), 6–10.

[10] Delmolino, K. et al. (2016). Step by step towards creating a safe smart contract: Lessons and insights from a cryptocurrency lab. In International Conference on Financial Cryptography and Data Security (pp. 79-94). Springer, Berlin, Heidelberg.

[11] Özyılmaz, K. R., & Yurdakul, A. (2021). Managing blockchain immutability in the GDPR era. Nature Electronics, 1-9.

[12] Hawlitschek, F., Notheisen, B., & Teubner, T. (2018). The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy. Electronic commerce research and applications, 29, 50-63.

[13] Zhang, P. et al. (2021). Blockchain technology in healthcare: Business models, use cases, challenges, and recommendations. Journal of Medical Internet Research, 23(8), e29106.

[14] Lin, I. C., & Liao, T. C. (2017). A survey of blockchain security issues and challenges. IJ Network Security, 19(5), 653-659.

[15] Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth, 10(1), 13-53.

[16] Zhang, A., Lin, X., & Sohal, A. S. (2020). Blockchain-based smart contracts: Applications, challenges, and future trends. Computer, 53(10), 50-57.

[17] Ramachandran, A., & Kantarcioglu, M. (2018). Using blockchain and smart contracts for secure data provenance management. arXiv preprint arXiv:1709.10000.

[18] Lin, J., Shen, Z., Miao, C., & Liu, A. X. (2018). Using blockchain technology to build trust in sharing LoRaWAN IoT data. In Proceedings of the 2nd International Conference on Crowd Science and Engineering (pp. 26-32).

[19] Mills, D. et al. (2016). Distributed ledger technology in payments, clearing, and settlement. Journal of Financial Market Infrastructures, 5(2), 215-238.

[20] Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., & Liu, Y. (2019). A survey on the scalability of blockchain systems. IEEE Network, 33(5), 166-173.

[21] Pahl, C., & El Ioini, N. (2020). Architectural principles for blockchain-based service interoperability. ACM Computing Surveys (CSUR), 54(5), 1-38.

[22] Alzahrani, N., & Bulusu, N. (2020). Blockchain Oracles: A Framework for Blockchain-Smart Contract Ecosystem. arXiv preprint arXiv:2004.07140.

[23] 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.