

A Literature Review on Data Monetization using Smart Contracts

Nishanth M¹ and Kavitha S²

Post Graduate Student, M.Tech Data Science, Department of Information Technology¹

Assistant Professor, M.Tech Data Science, Department of Information Technology²

Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India

Abstract: Blockchain technology is poised to change nearly every facet of our digital lives. Blockchain can be said as an immutable, and decentralized database. Data stored in the blockchain cannot be tampered, making it secure. Also, being decentralized, no central entity controls the blockchain, ensuring reliability. So, the data can be stored publicly, such that anyone could read the data. This vast availability of data could enable data scientists to perform various analytics over the large amount of data. This could result in many useful insights in many fields. But, when sensitive data such as healthcare data and reports are to be stored in the blockchain, it could raise several privacy issues. Medical reports or personal information cannot be stored in a way that anyone could access them. Thus, this paper suggests a way to store and perform analytics over sensitive data in blockchain. In this paper homomorphic encryption is used to store the sensitive data in blockchain. Computed results from homomorphic encryption on ciphertexts are encrypted. When the encrypted result is decoded, it produces a result that is identical to what would have happened if the operations had been carried out in plaintext. Thus, critical data are encrypted and stored in the blockchain, data analytics are performed over them, without knowing the actual data. Thus, the proposed system provides privacy of the data stored publicly and could also profit the data scientists with access to large amount of real time data directly from the owners of the data

Keywords: Blockchain technology, homomorphic encryption, health care data, data analytics

I. INTRODUCTION

For the past 20 years, the digital revolution has made data more and more accessible in ever larger quantities as analog processes have become digitized. Enterprises are under pressure to use personal information less freely, and more significantly, to evaluate it more carefully, as a result of new legislation like the GDPR. With "oldschool" tools, data scientists won't have as much access to data, but with the aid of Private Deep Learning, they will be able to analyse sensitive data without causing any privacy problems. When it comes to sensitive data such as in healthcare data of individuals cannot be collected and used. Healthcare data of a person cannot be exposed to others since it may cause several problems to the individual directly and indirectly. In this paper, We suggest a peer-to-peer network that allows anyone to train their AI models on private user information while retaining ownership and paying consumers for their data. This is achieved by using technologies like deep learning, Federated learning, homomorphic encryption and blockchain smart contracts

1.1 Data Monetization and Privacy– Need for the Study

In recent years, the proliferation of data and the advancements in technologies such as smart contracts, federated learning, and homomorphic encryption have opened up new opportunities and challenges in the field of data monetization and privacy. Understanding the need for study in this area is crucial for various stakeholders, including businesses, researchers, policymakers, and individuals. Data has become an invaluable asset for businesses across industries. By exploring data monetization strategies, organizations can unlock its potential and generate revenue streams. Smart contracts, which are self-executing contracts with predefined rules, offer a decentralized and transparent framework for facilitating data transactions. Studying the integration of smart contracts into data monetization processes can provide insights into new business models, revenue generation opportunities, and efficient data exchange

mechanisms. Privacy concerns have gained significant attention in the digital age. As data collection and sharing practices become more prevalent, individuals and organizations are increasingly concerned about their personal and sensitive information falling into the wrong hands. Federated learning, a distributed approach to machine learning, enables multiple parties to collaborate on model training without sharing raw data. Homomorphic encryption allows computation on encrypted data, preserving privacy while performing computations. Studying the application of federated learning and homomorphic encryption can help develop privacy-preserving mechanisms for data monetization, addressing privacy concerns and fostering trust between data providers and consumers. In conclusion, the need for study on the topic of data monetization using smart contracts and privacy using federated learning and homomorphic encryption arises from the increasing importance of data, privacy concerns, regulatory requirements, the quest for data utility, and the potential for future innovations. Understanding and addressing these aspects will help shape the future of data-driven economies while safeguarding individuals' privacy and data rights.

II. LITERATURE REVIEW

Kang, J., Wen, J., Ye, D., Lai, B., Wu, T., & Xiong, Z. (2023), "User-centric Incentive Mechanism with Optimal Data Freshness" proposes a novel approach to enhance the privacy and efficiency of healthcare data sharing in virtual healthcare environments. The authors suggest leveraging blockchain technology and federated learning to create a user-centric incentive mechanism that promotes active participation from users while ensuring optimal data freshness. By combining the benefits of blockchain's decentralized and immutable nature with federated learning's collaborative and privacy-preserving capabilities, the proposed framework enables healthcare data to be securely shared and analyzed while maintaining user privacy and control over their data. The incentive mechanism encourages user engagement and contributes to the overall freshness and quality of the shared data, ultimately enhancing the effectiveness of healthcare metaverses.

Neto, H. N. C., Hribar, J., Dusparic, I., Mattos, D. M., & Fernandes, N. C. (2023), The paper explores various applications of federated learning, analyzes potential attacks, highlights challenges, and presents emerging trends in securing this decentralized learning paradigm. It provides insights into the potential vulnerabilities and privacy risks associated with federated learning, including model inversion attacks and poisoning attacks. The analysis also addresses the challenges of ensuring data privacy, model integrity, and secure communication in federated learning systems. Furthermore, the paper discusses recent advancements and trends in securing federated learning, such as differential privacy techniques, secure aggregation protocols, and encryption methods. Overall, the study emphasizes the importance of robust security measures in federated learning to enable its widespread adoption while preserving data privacy and model integrity.

Shen, M., Gu, A., Kang, J., Tang, X., Lin, X., Zhu, L., & Niyato, D. (2023), The objective is to provide a comprehensive overview of the intersection between blockchain technology and the Artificial Intelligence of Things (AIoT). The survey aims to explore the potential benefits, challenges, and emerging trends in leveraging blockchain for secure and decentralized AIoT systems. It covers various aspects such as data integrity, trust, privacy, consensus mechanisms, smart contracts, and interoperability. The objective is to present a holistic understanding of how blockchain can enhance the security, reliability, and efficiency of AIoT deployments while ensuring transparency and accountability in the data exchange and decision-making processes.

Albulayhi, A. S., & Alsukayti, I. S. (2023), The objective of the paper is to propose an architecture that combines blockchain technology and the Internet of Things (IoT) for efficient management of IoT data communications through the use of smart contracts. The authors aim to address the challenges related to data security, trust, and privacy in IoT networks by leveraging blockchain's decentralized and tamper-resistant nature. By integrating smart contracts into the architecture, they seek to enhance automation, transparency, and reliability in IoT data communications, ultimately contributing to the development of more effective and secure IoT systems.

Su, Z., Guo, S., Dai, M., Luan, T. H., & Liu, Y. (2023), The objective of the survey on digital twins is to provide a comprehensive overview of the topic, covering various aspects including architecture, enabling technologies, security, privacy, and future prospects. The survey aims to explore the concept of digital twins, which refers to virtual representations of physical objects or systems, and their applications across different domains. It investigates the underlying architectures and technologies that enable the development and deployment of digital twins, while also

addressing the critical concerns of security and privacy. Furthermore, the survey offers insights into the potential future directions and prospects for digital twins, highlighting their potential impact and benefits in diverse industries and fields.

Nguyen, L. T., Nguyen, L. D., Hoang, T., Bandara, D., Wang, Q., Lu, Q., ... & Chen, S. (2023), The objective is to explore the potential of blockchain technology in enabling secure and reliable data sharing. The article aims to provide an overview of the fundamental concepts of blockchain and its application in establishing trust and privacy in data sharing scenarios. It also discusses the challenges associated with implementing blockchain-based solutions for data sharing, such as scalability, interoperability, and governance. Overall, the objective is to highlight the transformative capabilities of blockchain in ensuring trustworthy data sharing and to identify areas for further research and development.

Pithadia, H., Fenoglio, E., Batrinca, B., Treleaven, P., Echim, R., Bubutanu, A., & Kerrigan, C. (2023), The objective of this paper is to explore the concept of tokenizing data assets and developing methods to determine their value. By tokenizing data, it involves representing data as digital tokens on a blockchain or similar distributed ledger technology, enabling secure and transparent transactions. The goal is to create a framework that allows data to be treated as a valuable asset, similar to cryptocurrencies or other digital assets, and establish methods to assess its worth. This approach aims to provide a foundation for data monetization, where individuals or organizations can trade, sell, or exchange data tokens, leading to a more efficient and fair data economy.

Esmailzadeh, P. (2023), The objective of this paper is to explore the potential use of nonfungible tokens (NFTs) in facilitating the sharing of health information between different healthcare organizations. The author aims to examine how NFTs, which are unique digital assets that can represent ownership or access rights, can address the challenges of interoperability and privacy in health data exchange. By leveraging NFT technology, the article seeks to propose a novel approach to securely and efficiently share health information while ensuring data integrity, traceability, and patient privacy within the healthcare ecosystem.

Pabitha, P., Priya, J. C., Praveen, R., & Jagatheswari, S. (2023), To propose a novel blockchain framework called ModChain that addresses the security and scalability challenges in the Internet of Things (IoT) environment. The framework aims to enhance the security of IoT devices and transactions by incorporating hybrid techniques, including a combination of consensus mechanisms, data encryption, and access control. Additionally, ModChain focuses on improving the scalability of blockchain systems in IoT environments by employing a hierarchical structure and efficient data management techniques. The ultimate goal is to provide a secure and scalable blockchain framework that can effectively support the growing demands of IoT applications and enable trustworthy interactions among IoT devices.

Zirui, M., & Bin, G. (2023), The objective of the study is to develop a blockchain-based framework that ensures privacy preservation and user self-governance to address the issue of COVID-19-related depression in social media. The framework aims to provide individuals with a secure and confidential environment for sharing their experiences and emotions related to the pandemic, while also giving them control over their own data. By leveraging blockchain technology, the researchers aim to create a decentralized platform that promotes mental health support and community engagement while safeguarding user privacy and autonomy.

Dwivedi, S. K., Amin, R., & Vollala, S. (2023), The objective of the study conducted by Dwivedi, Amin, and Vollala (2023) is to propose a secure and trustworthy data storage and device authentication scheme in a fog computing environment. The scheme is based on smart contracts and the InterPlanetary File System (IPFS). The researchers aim to address the challenges of ensuring data security and device authentication in fog computing, which involves decentralized computing resources and data storage at the edge of the network. By leveraging smart contracts and IPFS, the proposed scheme aims to provide secure and reliable storage of data while enabling authentication of devices in a fog computing environment.

Ahsani, V., Rahimi, A., Letafati, M., & Khalaj, B. H. (2023), "The three pillars to watch: Privacy and Security, Edge Computing, and Blockchain" is to explore the key elements that are crucial for the successful implementation and development of Metaverse-as-a-Service. The authors focus on three main pillars: privacy and security, edge computing, and blockchain. The research aims to highlight the importance of these pillars and their potential role in ensuring the privacy, security, and efficient operation of Metaverse-as-a-Service platforms, which are virtual environments where users can interact with digital content and each other in real-time.

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023), The authors aim to provide insights into the evolving landscape of AI-driven data strategies, including the importance of data quality, data governance, and data integration techniques. The study seeks to contribute to the understanding of how organizations can effectively harness the potential of AI by adopting appropriate data strategies and integration approaches while addressing the existing challenges in this domain.

Treleven, P., Smietanka, M., & Pithadia, H. (2022), The article aims to provide an overview of federated learning, highlighting its potential to address privacy concerns by allowing data to remain decentralized and locally stored while enabling collaborative model training. It discusses the principles, challenges, and applications of federated learning, emphasizing its role in preserving data privacy and enabling large-scale collaborative learning without the need for centralized data repositories. The article seeks to shed light on the significance and future prospects of federated learning as a groundbreaking technology in the field of machine learning and data privacy.

Emish, M., Chaparala, H. K., Kelani, Z., & Young, S. D. (2022), The study focuses on the concept of data crowdsourcing, where individuals contribute their personal health data from wearable devices to a decentralized marketplace. The proposed marketplace utilizes blockchain technology to ensure data security, privacy, and transparency. Additionally, the research explores the application of federated machine learning, which allows for collaborative model training without sharing raw data, thereby addressing privacy concerns. Overall, the objective is to establish a framework that incentivizes individuals to share their wearable device data while ensuring data privacy and facilitating the development of machine learning models for healthcare advancements.

Yakubu, A. M., & Chen, Y. P. P. (2022), The objective of the study was to develop a blockchain-based application that utilizes smart contracts and homomorphic encryption for facilitating secure access to genomic data and discovering genetic variants. The researchers aimed to address the challenges of privacy and data security in genomics by leveraging the decentralized and transparent nature of blockchain technology, as well as the privacy-preserving capabilities of homomorphic encryption. The study proposed a novel approach to enable controlled access to genomic data while preserving privacy and confidentiality through the use of smart contracts and homomorphic encryption techniques.

Qammar, A., Karim, A., Ning, H., & Ding, J. (2022), The authors aim to address the privacy and security concerns associated with federated learning, which involves training machine learning models on decentralized data sources while preserving data privacy. By leveraging blockchain's distributed and immutable nature, the researchers propose a systematic framework that ensures the integrity, transparency, and trustworthiness of the federated learning process. The goal is to provide a secure and efficient environment for collaborative machine learning, where participants can contribute their data and computation resources without compromising privacy or data integrity.

Giaretta, L., Marchioro, T., Markatos, E., & Girdzijauskas, Š. (2022, December), The objective of the research presented in the paper "Towards a decentralized infrastructure for data marketplaces: narrowing the gap between academia and industry" is to bridge the divide between academic research and practical industry applications by developing a decentralized infrastructure for data marketplaces. The authors aim to address the challenges and limitations of centralized data marketplaces by proposing a decentralized approach that leverages blockchain technology and smart contracts. The research aims to contribute to the development of a more efficient, transparent, and secure framework for data exchange, promoting collaboration and innovation between academia and industry in the field of data economy.

Sun, J., Wu, Y., Wang, S., Fu, Y., & Chang, X. (2021), The objective of the study was to propose a permissioned blockchain framework that enhances the security of federated learning. Federated learning is a distributed machine learning approach that allows multiple parties to collaboratively train a model without sharing their private data. The researchers aimed to address the security and privacy concerns in federated learning by leveraging the immutability and transparency features of a permissioned blockchain. The proposed framework aimed to ensure the integrity and accountability of the model training process while preserving the privacy of the participants' data.

Jabarulla, M. Y., & Lee, H. N. (2021, August), The objective of the study is to propose a patient-centric healthcare system that combines blockchain technology and artificial intelligence (AI) to address the challenges posed by the COVID-19 pandemic. The study explores the opportunities and applications of this integrated approach, aiming to enhance healthcare delivery, patient data management, and decision-making processes. By leveraging blockchain's

secure and transparent nature along with AI's analytical capabilities, the proposed system seeks to improve the overall efficiency, effectiveness, and safety of healthcare services during the pandemic.

Śmietanka, M., Pithadia, H., & Treleaven, P. (2020), The objective of the paper "Federated Learning for Privacy-Preserving Data Access" by Śmietanka, Pithadia, and Treleaven is to explore the use of federated learning as a means of preserving privacy while enabling access to valuable data. The paper focuses on the challenges of traditional centralized data access models and proposes federated learning as a decentralized approach that allows multiple parties to collaborate on building machine learning models without sharing their raw data. The authors aim to highlight the potential of federated learning in addressing privacy concerns and providing a framework for secure and efficient data access in various domains.

Gupta, I. (2020), The objective of Gupta's paper, "Decentralization of Artificial Intelligence: Analyzing Developments in Decentralized Learning and Distributed AI Networks," is to explore and analyze the advancements and progress made in the field of decentralized learning and distributed AI networks.

The paper aims to provide insights into the potential benefits and challenges associated with decentralizing artificial intelligence systems, highlighting the emergence of new paradigms that distribute computation and knowledge across multiple nodes or devices. By examining the developments in this area, the objective is to contribute to a deeper understanding of the possibilities and implications of decentralization in artificial intelligence research and applications.

Ouyang, L., Yuan, Y., & Wang, F. Y. (2020), The objective of the research conducted by Ouyang, Yuan, and Wang is to propose a collaborative framework called "Learning Markets" that combines artificial intelligence (AI), blockchain technology, and smart contracts in the context of the Internet of Things (IoT). The framework aims to facilitate efficient and secure collaboration among AI agents and IoT devices, enabling data sharing, trading, and cooperative decision-making. By leveraging the transparency and decentralization features

of blockchain and the automation capabilities of smart contracts, the proposed framework aims to enhance trust, privacy, and efficiency in AI-enabled IoT environments, ultimately fostering more effective and scalable collaborations among various stakeholders.

Telenti, A., & Jiang, X. (2020), The objective of Telenti and Jiang's article titled "Treating medical data as a durable asset" is to highlight the value and potential of medical data as a long-term asset that can be used for research, clinical advancements, and personalized medicine. The authors argue that medical data, including genomic information, electronic health records, and other health-related data, should be considered as valuable resources that can be curated, shared, and leveraged over time to drive scientific discovery and improve patient outcomes. They emphasize the importance of establishing frameworks and infrastructures to ensure the secure and ethical use of medical data while promoting collaboration and data sharing among researchers and healthcare providers. The authors advocate for a shift in mindset, recognizing medical data as a durable asset that can significantly impact healthcare and the development of novel therapies.

Liu, Y., Yu, F. R., Li, X., Ji, H., & Leung, V. C. (2020), The objective of Liu et al.'s paper titled "Blockchain and Machine Learning for Communications and Networking Systems" is to provide a comprehensive survey and analysis of the potential applications, challenges, and integration strategies of blockchain and machine learning techniques in the field of communications and networking. The authors aim to explore how the combination of blockchain and machine learning can enhance the security, efficiency, and scalability of communication networks, and they discuss various use cases, methodologies, and algorithms that can leverage these technologies. The paper serves as a reference for researchers and practitioners interested in understanding the synergies between blockchain and machine learning in the context of communication systems.

III. CONCLUSION

As Blockchain is immutable, and a decentralized database, data stored in the blockchain cannot be tampered, making it secure. So, the data can be stored publicly, such that anyone could read the data which enables data scientists to perform various analytics over the large amount of data. This could result in many useful insights in many fields. In this paper way to store and perform analytics over sensitive data in blockchain is suggested with homomorphic encryption to enhance security of publicly stored data. Thus, critical data are encrypted and stored in the blockchain, data analytics are performed over them, without knowing the actual data. Thus, the proposed system provides privacy of the data

stored publicly and allows data scientists with access to large amount of real time data directly from the owners of the data.

REFERENCES

- [1]. Kang, J., Wen, J., Ye, D., Lai, B., Wu, T., & Xiong, Z. Blockchain-empowered Federated Learning for Healthcare Metaverses: User-centric Incentive Mechanism with Optimal Data Freshness.
- [2]. Neto, H. N. C., Hribar, J., Dusparic, I., Mattos, D. M., & Fernandes, N. C. (2023). Securing Federated Learning: A Security Analysis on Applications, Attacks, Challenges, and Trends. *IEEE Access*.
- [3]. Shen, M., Gu, A., Kang, J., Tang, X., Lin, X., Zhu, L., & Niyato, D. (2023). Blockchains for Artificial Intelligence of Things: A Comprehensive Survey. *IEEE Internet of Things Journal*.
- [4]. Albulayhi, A. S., & Alsukayti, I. S. (2023). A Blockchain-Centric IoT Architecture for Effective Smart Contract-Based Management of IoT Data Communications. *Electronics*, 12(12), 2564.
- [5]. Su, Z., Guo, S., Dai, M., Luan, T. H., & Liu, Y. (2023). A Survey on Digital Twins: Architecture, Enabling Technologies, Security and Privacy, and Future Prospects.
- [6]. Nguyen, L. T., Nguyen, L. D., Hoang, T., Bandara, D., Wang, Q., Lu, Q., ... & Chen, S. (2023). Blockchain-Empowered Trustworthy Data Sharing: Fundamentals, Applications, and Challenges. *arXiv preprint arXiv:2303.06546*.
- [7]. Pithadia, H., Fenoglio, E., Batrinca, B., Treleaven, P., Echim, R., Bubutanu, A., & Kerrigan, C. (2023). Data Assets: Tokenization and Valuation. Available at SSRN 4419590.
- [8]. Esmaeilzadeh, P. (2023). Evolution of Health Information Sharing Between Health Care Organizations: Potential of Nonfungible Tokens. *Interactive Journal of Medical Research*, 12(1), e42685.
- [9]. Pabitha, P., Priya, J. C., Praveen, R., & Jagatheswari, S. (2023). ModChain: a hybridized secure and scaling blockchain framework for IoT environment. *International Journal of Information Technology*, 15(3), 1741-1754.
- [10]. Zirui, M., & Bin, G. (2023). A privacy-preserved and user self-governance blockchain-based framework to combat COVID-19 depression in social media. *IEEE Access*.
- [11]. Dwivedi, S. K., Amin, R., & Vollala, S. (2023). Smart contract and ipfs-based trustworthy secure data storage and device authentication scheme in fog computing environment. *Peer-to-Peer Networking and Applications*, 16(1), 1-21.
- [12]. Ahsani, V., Rahimi, A., Letafati, M., & Khalaj, B. H. (2023). Unlocking Metaverse-as-a-Service The three pillars to watch: Privacy and Security, Edge Computing, and Blockchain. *arXiv preprint arXiv:2301.01221*.
- [13]. Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences*, 13(12), 7082.
- [14]. Treleaven, P., Smietanka, M., & Pithadia, H. (2022). Federated learning: the pioneering distributed machine learning and privacy-preserving data technology. *Computer*, 55(4), 20-29.
- [15]. Emish, M., Chaparala, H. K., Kelani, Z., & Young, S. D. (2022). On Monetizing Personal Wearable Devices Data: A Blockchain-based Marketplace for Data Crowdsourcing and Federated Machine Learning in Healthcare. *Artificial Intelligence Advances*, 4(2).
- [16]. Yakubu, A. M., & Chen, Y. P. P. (2022). A blockchain-based application for genomic access and variant discovery using smart contracts and homomorphic encryption. *Future Generation Computer Systems*, 137, 234-247.
- [17]. Qammar, A., Karim, A., Ning, H., & Ding, J. (2022). Securing federated learning with blockchain: a systematic.
- [18]. Giaretta, L., Marchioro, T., Markatos, E., & Girdzijauskas, Š. (2022, December). Towards a decentralized infrastructure for data marketplaces: narrowing the gap between academia and industry. In *Proceedings of the 1st International Workshop on Data Economy* (pp. 49-56).
- [19]. Sun, J., Wu, Y., Wang, S., Fu, Y., & Chang, X. (2021). Permissioned blockchain frame for secure federated learning. *IEEE Communications Letters*, 26(1), 13-17.

- [20]. Jabarulla, M. Y., & Lee, H. N. (2021, August). A blockchain and artificial intelligence-based, patient-centric healthcare system for combating the COVID-19 pandemic: Opportunities and applications. In *Healthcare* (Vol. 9, No. 8, p. 1019). MDPI.
- [21]. Śmietanka, M., Pithadia, H., & Treleaven, P. (2020). Federated learning for privacy-preserving data access. Available at SSRN 3696609.
- [22]. Gupta, I. (2020). Decentralization of artificial intelligence: analyzing developments in decentralized learning and distributed AI networks. arXiv preprint arXiv:1603.04467.
- [23]. Ouyang, L., Yuan, Y., & Wang, F. Y. (2020). Learning markets: An AI collaboration framework based on blockchain and smart contracts. *IEEE Internet of Things Journal*, 9(16), 14273-14286.
- [24]. Telenti, A., & Jiang, X. (2020). Treating medical data as a durable asset. *Nature Genetics*, 52(10), 1005-1010.
- [25]. Liu, Y., Yu, F. R., Li, X., Ji, H., & Leung, V. C. (2020). Blockchain and machine learning for communications and networking systems. *IEEE Communications Surveys & Tutorials*, 22(2), 1392-1431.