International Journal of Computer Sciences and Engineering
Vol.11, Issue.12, pp.32-37, December 2023
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

IJCSE
ISSN: 2347-2693 (E)

Review Paper

Blockchain Based Data Privacy through Artificial Intelligence: Review

Shashank Saroop[1*], Radha[2]

[1,2]Dept. of CSE, MIET Greater Noida, India

**Corresponding Author: shashank.saroop@mietengineering.org*

Received: 30/Oct/2023; **Accepted**: 02/Dec/2023; **Published**: 31/Dec/2023. **DOI**: https://doi.org/10.26438/ijcse/v11i12.3237

Abstract: Data privacy and security have become paramount concerns in the realm of artificial intelligence (AI) due to the increasing reliance on vast datasets for training AI models. This review paper explores the potential of blockchain technology to enhance data privacy and security in AI applications. Blockchain, known for its core features of decentralization, immutability, transparency, and security, offers a promising framework to address data privacy challenges. Keywords like decentralized data storage, access control mechanisms, data provenance, and privacy-preserving machine learning are discussed in the context of blockchain integration with AI. Several use cases, including healthcare, finance, supply chain, and identity verification, demonstrate the practical applicability of blockchain in safeguarding sensitive data. However, challenges related to scalability, regulation, and adoption must be addressed. The paper concludes by highlighting emerging trends, research directions, and the importance of ongoing efforts to harness blockchain's potential for preserving data privacy in AI.

Keywords: Blockchain, Data Privacy, Artificial Intelligence, Decentralized Data Storage, Access Control, Data Provenance, Privacy-Preserving Machine Learning, Use Cases, Challenges, Emerging Trends.

1. Introduction

Data privacy has emerged as a critical concern in the field of artificial intelligence (AI) due to the growing reliance on vast and sensitive datasets for training machine learning models. In the age of data-driven decision-making, AI systems are increasingly employed in various domains, including healthcare, finance, supply chain management, and more. These AI systems often require access to extensive and highly personal data, such as medical records, financial transactions, and personal identification information. Consequently, ensuring the privacy and security of this data is of utmost importance.

Data privacy in AI not only safeguards individuals' personal information but also plays a pivotal role in building trust in AI technologies. Users, organizations, and governments are becoming increasingly concerned about the potential misuse or mishandling of data, leading to privacy breaches and unauthorized access. Therefore, it is imperative to develop robust mechanisms that protect the privacy of data used in AI applications while still enabling the advancement of AI technologies.

1.1. Statement of the Problem: Challenges in Maintaining Data Privacy in AI

The challenges in maintaining data privacy in AI applications are multifaceted. Firstly, the collection and storage of large datasets required for training AI models can make individuals vulnerable to data breaches and privacy violations. High-profile incidents, such as data leaks and unauthorized data access, have raised significant concerns about the security of centralized data repositories.

Secondly, as AI systems become more complex and distributed, ensuring that data remains private throughout its lifecycle, from collection to analysis, becomes increasingly challenging. The need to share data among various stakeholders, such as data providers, model developers, and end-users, poses additional privacy risks. Furthermore, regulations and legal frameworks (e.g., GDPR, HIPAA) have been enacted to protect individuals' data privacy rights. Non-compliance with these regulations can result in substantial fines and damage to an organization's reputation. Hence, organizations must navigate the intricate landscape of data privacy laws while deploying AI solutions.

1.2. Purpose of the Review: Exploring Blockchain as a Solution

The purpose of this review paper is to explore the potential of blockchain technology as a solution to address the challenges of maintaining data privacy in AI applications. Blockchain, originally designed as a decentralized and immutable ledger for cryptocurrencies like Bitcoin, has evolved into a versatile technology with attributes that align well with data privacy requirements.

Blockchain offers decentralization, which eliminates single points of failure and reduces the risk of unauthorized access

to data. Its immutability ensures the integrity of stored data, and transparency enables auditability. Additionally, blockchain can facilitate secure data sharing and access control through smart contracts. This paper aims to provide a comprehensive overview of blockchain's role in enhancing data privacy and security in AI applications. It will delve into various aspects of blockchain technology, its potential use cases, challenges, and emerging trends in the intersection of blockchain and AI data privacy.

2. Background

2.1. Definition of Blockchain Technology

Blockchain technology is a decentralized and distributed ledger system originally designed to support cryptocurrencies like Bitcoin [1]. It consists of a chain of blocks, where each block contains a set of transactions or data. These blocks are linked together in chronological order, creating a secure and tamper-resistant ledger.

At its core, blockchain relies on several key principles:

- Decentralization: Data is not stored in a central authority or server but is distributed across a network of nodes, making it resistant to single points of failure.
- Immutability: Once data is recorded in a block, it cannot be altered or deleted, ensuring the integrity of the ledger.
- Transparency: The ledger is publicly accessible, allowing participants to view transactions and data in a transparent manner.
- Security: Cryptographic techniques secure data and transactions, making it extremely difficult for unauthorized parties to manipulate the blockchain.

2.2. Overview of AI Applications and Their Data Requirements

AI applications encompass a wide range of domains, including natural language processing, computer vision, recommendation systems, and more. These applications rely heavily on data, specifically labeled datasets, for training and improving models. The larger and more diverse the dataset, the better AI models can perform [2]. For instance, in computer vision, deep learning models require extensive image datasets for tasks like object detection and facial recognition. Similarly, in healthcare, AI models may need access to electronic health records (EHRs) to diagnose diseases or predict patient outcomes.

2.3. Data Privacy Concerns in AI

The collection and use of large, sensitive datasets in AI applications raise significant data privacy concerns. These concerns include:

- Data Breaches: The storage of massive datasets in centralized repositories makes them attractive targets for hackers and malicious actors. High-profile data breaches, such as the Equifax breach in 2017 [3], have exposed the personal information of millions of individuals.
- Unauthorized Access: Data used in AI can be accessed and used for purposes beyond its original intent, potentially infringing on individuals' privacy rights.
- Data Bias: AI models trained on biased or incomplete datasets may perpetuate and amplify existing biases, leading to discriminatory outcomes [4].
- Regulatory Compliance: Organizations must comply with data privacy regulations like the European Union's General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) in the United States, adding complexity to AI development and deployment [5].

2.4. Existing Solutions and Their Limitations

Several existing solutions attempt to address data privacy concerns in AI, including data anonymization, differential privacy, and federated learning. Data anonymization involves removing or obfuscating personally identifiable information from datasets. Differential privacy adds noise to query results to protect individuals' privacy. Federated learning allows AI models to be trained across decentralized data sources without centralizing the data. However, these solutions have limitations. Anonymization techniques can be reversed, exposing sensitive information [6]. Differential privacy may impact the utility of data for analysis. Federated learning introduces communication overhead and coordination challenges [7].

3. Blockchain Technology

3.1. Types of Blockchains

Public Blockchain:

Public blockchains are open to anyone and are permissionless. Participants can join the network, validate transactions, and create blocks without needing authorization. Examples include Bitcoin and Ethereum. Public blockchains offer high levels of decentralization and transparency.

Private Blockchain:

Private blockchains are restricted to a specific group of participants who have permission to access and validate the network. These networks are often used by enterprises for internal purposes, providing greater control over data privacy and security [8].

Consortium Blockchain:

Consortium blockchains are a hybrid approach, where a group of organizations collaboratively manages the network. Consortium members have specific permissions to validate transactions. This type of blockchain strikes a balance between public and private blockchains and is suitable for industries requiring shared data governance [9].

3.2. Smart Contracts and Their Role in Data Access Control

Smart contracts are self-executing, programmable contracts that run on blockchain networks [10]. They play a crucial role in data access control by automating and enforcing predefined rules and conditions.

Role-Based Access Control: Smart contracts can define access permissions based on roles within an organization. For

example, in a healthcare blockchain, a smart contract can specify that only authorized healthcare providers can access specific patient records [11].

- Data Sharing Agreements: Smart contracts enable parties to establish and enforce data-sharing agreements. These contracts can specify the terms and conditions under which data can be accessed or shared, ensuring transparency and accountability [12].
- Data Provenance: Smart contracts can track the history of data access and modifications. This feature enhances transparency and audibility, crucial for maintaining data privacy [13].
- Conditional Access: Smart contracts can enforce access conditions. For example, data access may be granted only after the fulfillment of certain conditions, such as consent from data owners or compliance with regulatory requirements [14].

4. Blockchain for Data Privacy in AI

4.1. Decentralized Data Storage

- Storing AI Training Datasets on a Blockchain:

Blockchain technology allows for the storage of AI training datasets in a decentralized manner. Training data can be hashed, encrypted, and distributed across the nodes of the blockchain network [14]. This ensures that the data is not stored in a single central repository but is instead distributed, reducing the risk of data breaches.

- Benefits and Challenges:

Decentralized data storage on a blockchain enhances data security by reducing the vulnerability associated with centralized data repositories. It also provides a transparent and auditable record of data transactions.

Challenges include the scalability of blockchain networks for storing large datasets, the cost associated with data storage on the blockchain, and the need to ensure data privacy while maintaining transparency [15].

4.2. Access Control Mechanisms

Role of Smart Contracts in Controlling Data Access:

Smart contracts play a crucial role in managing data access in blockchain-based AI systems. They can define access permissions and conditions, enabling fine-grained control over who can access and modify data [16].

Case Studies of Blockchain-Based Access Control in AI:

Several case studies demonstrate the use of smart contracts for access control in AI. For instance, a healthcare consortium may employ smart contracts to govern access to patient records, ensuring that only authorized healthcare providers can access and update the data [17].

Data Provenance and Auditability

How Blockchain Can Provide Data Traceability:

Blockchain's transparent and immutable ledger ensures data traceability. Every data transaction, including who accessed it and when, is recorded in the blockchain. This feature enables auditing and accountability in AI systems [17].

Ensuring Data Integrity in AI Models:

Blockchain can be used to verify the integrity of AI models. By recording the training process and dataset details on the blockchain, users can ensure that AI models are trained on authentic and unaltered data [17].

Privacy-Preserving Machine Learning on Blockchain

Techniques for Training AI Models while Preserving Data Privacy:

Privacy-preserving machine learning techniques can be integrated with blockchain to protect sensitive data during model training. Techniques such as federated learning and secure multi-party computation enable AI models to be trained without exposing raw data [18].

Homomorphic Encryption, Federated Learning, etc.:

Homomorphic encryption allows computations to be performed on encrypted data, preserving privacy. Federated learning allows models to be trained on decentralized data sources without sharing the data itself [18].

5. Use Cases and Applications

5.1. Healthcare

- *Securing Patient Data for Medical AI Applications:*

In the healthcare sector, patient data privacy is of utmost importance. Blockchain can be utilized to secure patient data used in medical AI application [19]. Patient records, medical images, and other sensitive health data can be stored on a blockchain with stringent access control and auditability through smart contracts. This ensures that only authorized healthcare professionals can access patient information, enhancing data privacy [19].

5.2. Finance

- *Fraud Detection and Transaction Monitoring:*

Blockchain technology can enhance data privacy and security in financial applications, particularly in fraud detection and transaction monitoring [20]. Financial institutions can use blockchain to securely share transaction data and detect fraudulent activities while preserving the confidentiality of sensitive customer information. The transparency and immutability of blockchain transactions make it easier to trace and verify financial transactions [21].

5.3. Supply Chain

- *Ensuring Transparency and Authenticity in Supply Chain AI:*

Blockchain plays a crucial role in ensuring transparency and authenticity in supply chain management. In supply chain AI applications, blockchain can be used to record the provenance of goods, track their journey from manufacturer to consumer, and verify the authenticity of products (21). This not only enhances data privacy but also helps in reducing counterfeiting and fraud.

5.4. Identity Verification

- *Using Blockchain for Secure Identity Verification in AI Systems:*

Blockchain technology can be leveraged for secure identity verification in various AI systems. For example, in online identity verification processes, individuals' identity credentials can be stored on a blockchain, and access to this data can be controlled through smart contracts [22]. Users can grant temporary and controlled access to their identity information, improving privacy and security in online transactions.

6. Challenges and Limitations

6.1. Scalability and Performance Issues

Scalability: One of the primary challenges of blockchain is its scalability. Public blockchains like Bitcoin and Ethereum have faced limitations in handling a high volume of transactions and data storage [22]. The addition of data, especially large AI training datasets, can strain the network, resulting in slower transaction processing times and increased costs (22].

Performance: As the blockchain network grows, the performance can degrade, impacting the efficiency of data access and verification. Slow transaction confirmation times and high transaction fees on public blockchains may hinder the feasibility of using blockchain for data-intensive AI applications (22].

6.2. Regulatory and Compliance Challenges

Data Privacy Regulations: Compliance with data privacy regulations such as the General Data Protection Regulation (GDPR) in the European Union and the Health Insurance Portability and Accountability Act (HIPAA) in the United States poses significant challenges [23]. Blockchain's transparency can conflict with GDPR's "right to be forgotten" and HIPAA's strict healthcare data handling requirements.

Legal Uncertainty: The legal status of blockchain and smart contracts varies across jurisdictions, leading to legal uncertainty [23]. This can hinder the adoption of blockchain for data privacy in AI as organizations grapple with compliance.

6.3. Adoption and Interoperability Concerns

Interoperability: The lack of standardized protocols and interoperability among different blockchain platforms and networks hinders the seamless exchange of data [23]. This can complicate efforts to integrate blockchain into existing AI systems and workflows.

Adoption Barriers: Organizations may face resistance in adopting blockchain technology due to the complexity of implementation, lack of in-house expertise, and concerns about the technology's immaturity [23]. These adoption barriers can slow down the deployment of blockchain-based data privacy solutions in AI.

6.4. Environmental Impact of Blockchain Networks

Energy Consumption: Proof of Work (PoW) consensus mechanisms, used in prominent blockchains like Bitcoin and Ethereum, are known for their high energy consumption [24]. The environmental impact of running blockchain networks, particularly those with large numbers of miners, can be substantial.

Sustainability Concerns: The environmental concerns related to blockchain's energy consumption can clash with sustainability goals, leading to ethical dilemmas for organizations that aim to reduce their carbon footprint [24].

7. Future Directions

7.1. Emerging Trends and Innovations in Blockchain for AI Data Privacy

Zero-Knowledge Proofs: Zero-knowledge proofs, such as zk-SNARKs, enable data validation without revealing the data itself. This technology has the potential to enhance privacy-preserving AI on the blockchain [24].

Layer 2 Solutions: Layer 2 scaling solutions, like Lightning Network for Bitcoin and Layer 2 protocols for Ethereum, aim to improve blockchain scalability and reduce transaction fees. These advancements could make blockchain more practical for AI data storage and processing.

Sidechains and Cross-Chain Integration: Efforts to connect multiple blockchains through sidechains and interoperability protocols could enable seamless data sharing and access control across different blockchain networks.

7.2. Research Areas that Need Further Exploration

Privacy-Preserving Consensus Mechanisms: Developing consensus algorithms that prioritize privacy while maintaining security and scalability is an ongoing research challenge.

Standardization and Interoperability: More research is needed to standardize data formats and interoperability protocols to facilitate the exchange of data and smart contracts across different blockchain networks [25].

Usability and User Education: Research should focus on making blockchain-based AI data privacy solutions more user-friendly and educating users on the technology's benefits and risks [25].

7.3. Potential Integration with Other Privacy-Preserving Technologies

Homomorphic Encryption: Combining blockchain with homomorphic encryption allows computations on encrypted data, enabling AI model training without exposing raw data.

Federated Learning: Integrating blockchain with federated learning enables privacy-preserving model training on decentralized data sources. This approach can be particularly useful in sectors like healthcare.

Multi-Party Computation (MPC): Combining MPC with blockchain allows multiple parties to jointly compute AI models without revealing their individual data. This approach offers strong privacy guarantees.

Differential Privacy: Blockchain-based systems can incorporate differential privacy techniques to protect individual privacy while aggregating AI insights from distributed data

Secure Enclaves: Leveraging hardware-based secure enclaves can enhance data privacy within blockchain networks by isolating sensitive computations.

8. Conclusion

In recent years, the intersection of blockchain technology and artificial intelligence (AI) has shown great promise in addressing data privacy and security concerns in AI applications. This review paper has explored the importance of data privacy in AI, the challenges it poses, and how blockchain can serve as a solution. We have discussed the core principles of blockchain, its types, and the role of smart contracts in data access control. Additionally, we have examined use cases across industries, including healthcare, finance, supply chain, and identity verification, where blockchain can enhance data privacy. However, while blockchain offers significant advantages in enhancing data privacy in AI, it is important to acknowledge the challenges and limitations associated with its implementation. Scalability and performance issues in public blockchains, regulatory and compliance challenges, adoption barriers, and environmental concerns need to be carefully addressed. Looking ahead, several promising trends and innovations are emerging in the field of blockchain for AI data privacy. Zero-knowledge proofs, layer 2 scaling solutions, and cross-chain integration are set to improve the scalability and efficiency of blockchain networks. Research areas such as privacy-preserving consensus mechanisms, standardization, usability, and legal frameworks are ripe for further exploration. Blockchain is not an isolated solution; it can be integrated with other privacy-preserving technologies like homomorphic encryption, federated learning, multi-party computation (MPC), differential privacy, and secure enclaves to create comprehensive privacy solutions for AI applications. In conclusion, blockchain technology offers a compelling avenue to enhance data privacy and security in AI. While challenges exist, ongoing research and innovation hold the potential to overcome these hurdles. The fusion of blockchain and AI represents a promising frontier in the pursuit of privacy-preserving AI systems that empower individuals and organizations to securely leverage data for innovation and progress. As the field continues to evolve, it is imperative for researchers, practitioners, and policymakers to collaborate and shape the future of blockchain for AI data privacy, ensuring a balance between innovation, security, and ethical considerations.

References

[1]. Sweeney, L. K-anonymity: "A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems" Vol.**10**, Issue.**5**, pp.**557-570, 2002.**

[2]. Dwork, C, McSherry, F, Nissim, K, & Smith, A. "Calibrating noise to sensitivity in private data analysis". In Proceedings of the Third Conference on Theory of Cryptography (TCC'06), **2002.**

[3]. Nakamoto, S. "Bitcoin: A Peer-to-Peer Electronic Cash System" https://bitcoin.org/bitcoin.pdf, **2008.**

[4]. Lauter, K., Naehrig, M., & Vaikuntanathan, V. "Can homomorphic encryption be practical" In Proceedings of the 3rd ACM workshop on Cloud computing security workshop (CCSW'11), **2008.**

[5]. Back, A, Corallo, M., Dashjr, L., Friedenbach, M., Maxwell, G., Miller, A., ... & Wuille, P." Enabling Blockchain Innovations with Pegged Sidechains", **2014.**

[6]. Ben-Sasson, E., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., ... & Virza, M. (2014). Zero Cash: Decentralized anonymous payments from Bitcoin in IEEE Symposium on Security and Privacy, **2014.**

[7]. LeCun, Y. Bengio, Y. & Hinton, G. "Deep learning. Nature", 521(7553), pp.**436-444, 2014.**

[8]. Swan, M. "Blockchain: Blueprint for a New Economy" O'Reilly Media, **2015.**

[9]. Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J. A., & Felten, E. W. "Research perspectives and challenges for bitcoin and cryptocurrencies" in Proceedings of the IEEE Symposium on Security and Privacy. **2015.**

[10]. Intel. (2016). Intel Software Guard Extensions (Intel SGX) for Fun and Profit. **2016.**

[11]. O'Neil, C. "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy Crown", **2016.**

[12]. Poon, J., & Dryja, T "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments" https://lightning.network/lightning-network-paper.pdf . **2016.**

[13]. Mougayar, W. "The Business Blockchain: Promise, Practice, and Application of the Next Internet Technology", Wiley. **2016**

[14]. Tapscott, D., & Tapscott, A. "Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World. Penguin" **2016.**

[15]. Kosba, A., Miller, A., Shi, E., Wen, Z., & Papamanthou, C. Hawk "The blockchain model of cryptography and privacy-preserving smart contracts" in Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS'16), **2016**

[16]. Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. "Blockchain technology: Beyond bitcoin Applied Innovation", 6-10, pp.**71-81, 2016.**

[17]. Voigt. P, Von dem Bussche, A. "The EU General Data Protection Regulation" (GDPR): A Practical Guide. Springer. **2017.**

[18]. B., Patel, S., ... & Viswanathan, V. "Practical secure aggregation for privacy-preserving machine learning". In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS'17), **2017.**

[19]. McMahan, H. B. Moore, E. Ramage, D. Hampson, S. & Arcas, "Communication-efficient learning of deep networks from decentralized data", in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), **2018.**

[20]. Gupta, A. Capretz, L. F. & Ahmed, F. "Towards understanding blockchain roles in improving the Internet of Things". in IEEE 5th World Forum on Internet of Things (WF-IoT) (pp. 127-132). **2018.**

[21]. Mengelkamp. E, Notheisen, B, Bräuer, S, & Flath, C. M. "A blockchain-based smart grid: towards sustainable local energy markets" in Computer Science-Research and Development, Vol.**33**, Issue.1-2, pp.**207-214**, **2018.**

[22]. Al-Bassam. M, Sonnino. A, Bano. S, Hrycyszyn, D., Danezis, G., & Papamanthou, C. (2018). Chainspace: A Sharded Smart Contracts Platform. In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS'18). **2018.**

[23]. Hernandez, C. (2019). 6 "Biggest Data Breaches in History Forbes. www.forbes.com/sites/christopherhelman2019/01/19/6-of-the biggest-data-breaches-in-history. **2019.**

[24]. Kairouz. P, McMaha. B, Avent. B, Bellet. A. Bennis. M, Bhagoji. Zhang, S. "Advances and open problems in federated learning". preprint arXiv:1912.04, **2019.**

AUTHORS PROFILES.

Shashank Saroop earned his B. Tech., from UPTU, in 2009 M. Tech., from NSUT (East Campus) in 2011 and PhD Pursuing from Amity University, Gurugram in Computer Science & Engineering. He is currently working as Assistant Professor in Department of Computer Science & Engineering in MIET, Greater Noida since 2023. He has published more than 10 research papers in reputed international journals including (SCI & Web of Science) and conferences including IEEE . His main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. He has 12 years of teaching experience.

Radha earned his B. Tech., from MDU, in 2008 M. Tech., from MDU in 2013 and PhD Pursuing in Computer Science & Engineering He is currently working as Assistant Professor in Department of Computer Science & Engineering in MIET, Greater Noida since 2023. He has published more than 5 research papers in reputed international journals and conferences including IEEE . His main research work focuses on Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT He has 12 years of teaching experience.

