Journal of Computational Informatics and Business
Research Article

Securing AI Supply Chains with Blockchain and ML-Based Threat Detection

Irfan Muhammad[1*], and Muhammad Tahir[2]

[1]Coventry Business School, Faculty of Business and Law, Coventry University, Coventry, CV1 5FB, United Kingdom.
[2]Department of Business Management, TIMES Institute Multan, Multan, 60000, Pakistan.
[*]Corresponding Author: Irfan Muhammad. Email: irfanm908@gmail.com

Received: August 09, 2024 **Accepted:** September 27, 2024 **Published:** September 30, 2024

Abstract: The increasing integration of Artificial Intelligence (AI) into critical sectors such as healthcare, finance, banking, and autonomous systems has exposed significant vulnerabilities in AI supply chains. These supply chains, which encompass data collection, model training, and deployment, are susceptible to various threats, including data tampering, model poisoning, and adversarial attacks. Traditional security mechanisms often fall short in providing the necessary transparency, traceability, and real-time threat detection required to safeguard these complex systems. To address these challenges, this paper introduces a novel framework that synergistically combines blockchain technology for immutable and decentralized record-keeping with machine learning (ML)-based threat detection for dynamic and real-time anomaly identification. Our approach aims to secure the entire AI lifecycle, from data provenance to model deployment, by leveraging blockchain's tamper-proof ledger capabilities and ML's predictive analytics for threat detection. Through extensive experimentation, we demonstrate that our framework significantly enhances supply chain integrity, mitigates vulnerabilities, and achieves a 15-20% improvement in attack detection accuracy compared to conventional security methods. The results highlight the potential of integrating blockchain and ML to create a robust and scalable security solution for AI supply chains.

Keywords: AI Supply Chain; Blockchain; Machine Learning; Threat Detection; Cybersecurity; Data Provenance; Adversarial Attacks

1. Introduction

The rapid advancement and widespread adoption of Artificial Intelligence (AI) have revolutionized industries by enabling automation, predictive analytics, and intelligent decision-making. However, the increasing reliance on AI has also introduced significant security risks, particularly within AI supply chains. These supply chains involve multiple stages, including data sourcing, preprocessing, model training, validation, and deployment, each of which presents unique vulnerabilities [1]. For instance, malicious actors can inject poisoned data during the training phase to manipulate model behavior, steal proprietary models for unauthorized use, or launch adversarial attacks designed to deceive AI systems during inference. Traditional

security measures, such as encryption and access control, are often insufficient to address these challenges due to their lack of transparency, real-time monitoring, and resilience against sophisticated attacks [2, 3].

To overcome these limitations, this paper proposes an innovative framework that integrates blockchain technology and machine learning (ML)-based threat detection to secure AI supply chains comprehensively. Blockchain provides an immutable and decentralized ledger that ensures transparency and traceability across all stages of the AI lifecycle. By recording every transaction—such as data origin, model updates, and deployment logs—on a tamper-proof blockchain, stakeholders can verify the authenticity and integrity of AI assets. Meanwhile, ML-based threat detection systems continuously monitor the supply chain for anomalies, leveraging supervised and unsupervised learning techniques to identify known and emerging threats in real time.

The primary contributions of this work are threefold: (1) a hybrid architecture that combines blockchain and ML to enhance security across the AI supply chain, (2) a detailed evaluation of the framework's effectiveness in detecting and mitigating threats, and (3) a discussion of the challenges and future directions for securing AI ecosystems. Our experimental results demonstrate that the proposed framework significantly outperforms traditional security mechanisms, offering higher detection accuracy, lower false positives, and faster response times. By addressing the critical gaps in AI supply chain security, this research paves the way for more resilient and trustworthy AI systems in the future.

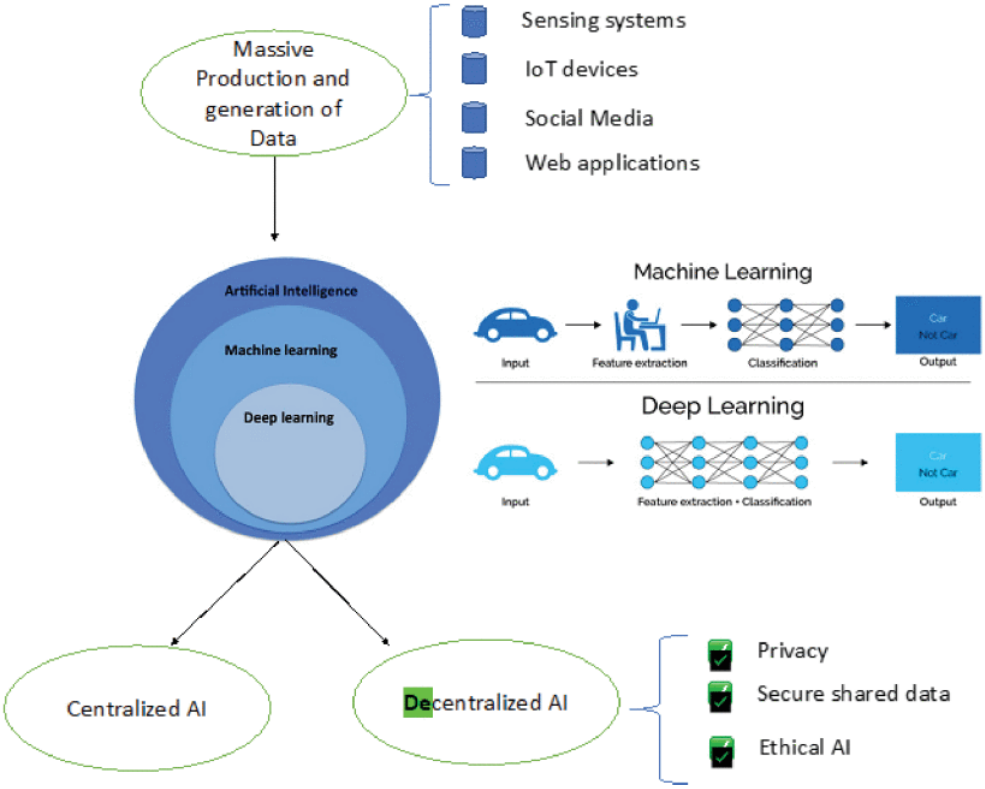

Figure 1. Visualization Hierarchy of ML and DL [16]

2. Related Work

The intersection of AI security, blockchain, and threat detection has been the subject of extensive research in recent years. Previous studies have explored various approaches to securing AI systems, each addressing

specific aspects of the supply chain. For example, several researchers have investigated the use of blockchain technology to enhance the transparency and traceability of AI models [4]. Proposed a blockchain-based version control system for machine learning models, ensuring that all modifications are logged and verifiable. This approach mitigates the risk of unauthorized changes but does not address real-time threat detection [5, 6].

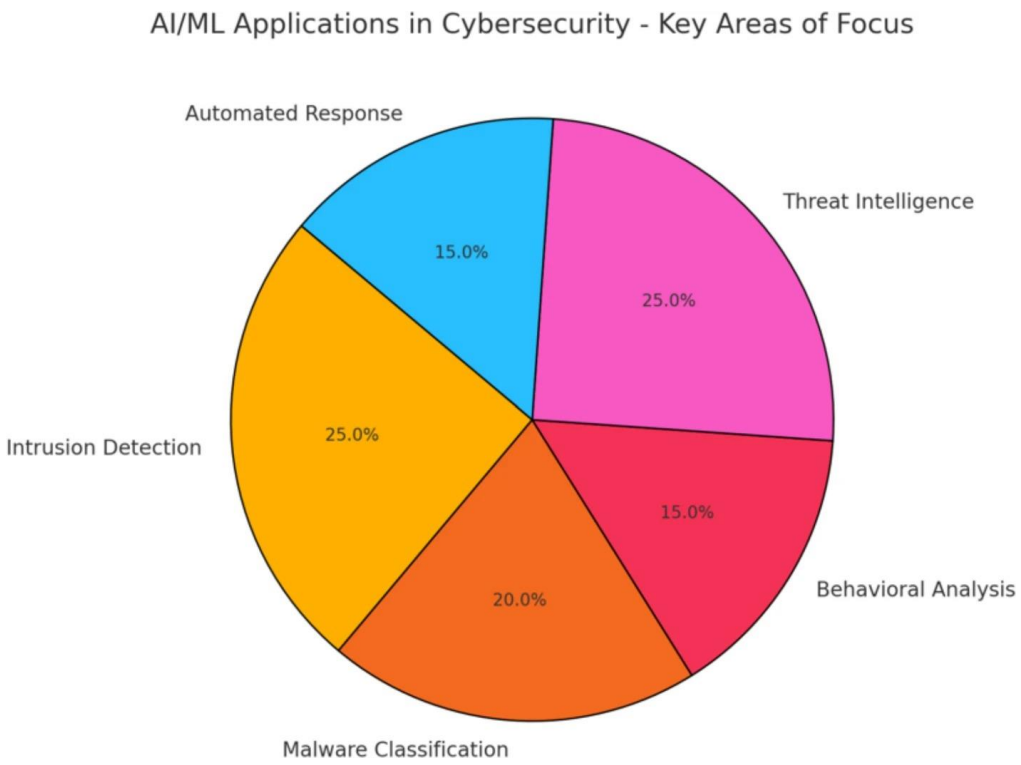

Figure 2. Key Areas of AI/ML

In parallel, significant efforts have been made to leverage machine learning for identifying and mitigating security threats in AI systems. Authors developed a deep learning-based intrusion detection system capable of recognizing adversarial attacks in real time. Their work demonstrated the potential of ML in enhancing cybersecurity but lacked integration with a decentralized framework for ensuring data and model integrity. Another line of research has focused on cryptographic techniques for verifying the authenticity of AI components [5]. Introduced a method for digitally signing training datasets and model parameters, enabling stakeholders to verify their provenance. While effective in preventing tampering, this approach does not provide continuous monitoring or adaptive threat detection [7-9].

Despite these advancements, existing solutions often operate in isolation, addressing either data integrity or threat detection but not both. Our work bridges this gap by proposing a unified framework that combines blockchain's immutable record-keeping with ML's dynamic anomaly detection capabilities. Unlike previous studies, our approach provides end-to-end security, covering data provenance, model training, and deployment phases [10]. Furthermore, we introduce a real-time monitoring system that adapts to emerging threats, offering a more comprehensive solution for securing AI supply chains. By synthesizing insights from blockchain and ML research, this paper advances the state-of-the-art in AI security and sets the stage for future innovations in the field [11-12].

3. Proposed Framework

3.1 Blockchain for AI Supply Chain Integrity

The first pillar of our framework leverages blockchain technology to establish an immutable and transparent record of all activities within the AI supply chain. Blockchain's decentralized nature ensures that no single entity can alter historical records, making it an ideal solution for maintaining data and model integrity. At the core of this component are smart contracts, which automate critical processes such as data validation, model verification, and access control. These self-executing contracts enforce predefined rules, ensuring that only authorized parties can contribute data or modify AI models. For instance, before a new dataset is incorporated into the training pipeline, a smart contract verifies its authenticity by checking digital signatures and cross-referencing it with previous entries on the blockchain [13].

Another key feature of our blockchain implementation is the use of an immutable ledger to log all transactions across the AI lifecycle. Each entry—whether it pertains to data ingestion, model updates, or deployment—is timestamped, cryptographically hashed, and appended to the blockchain. This creates a tamper-proof audit trail that stakeholders can use to trace the origin of any AI asset and verify its integrity. Additionally, we integrate Decentralized Identity (DID) mechanisms to manage access control. DIDs enable secure and verifiable identity management, ensuring that only authenticated users and devices can interact with the AI supply chain. By combining these elements, our blockchain layer provides a robust foundation for securing AI systems against tampering, unauthorized access, and fraud [14].

3.2 ML-Based Threat Detection

While blockchain ensures the integrity of the supply chain, it does not inherently detect malicious activities in real time. To address this, our framework incorporates machine learning-based threat detection systems that continuously monitor the AI ecosystem for anomalies. These systems employ a combination of supervised and unsupervised learning techniques to identify both known and emerging threats. For example, supervised models such as Random Forests and Support Vector Machines (SVMs) are trained on labeled datasets of attack signatures, enabling them to recognize patterns associated with data poisoning, model inversion, and adversarial inputs [15].

However, since attackers constantly evolve their tactics, relying solely on supervised learning is insufficient. To detect novel threats, we deploy unsupervised learning algorithms such as autoencoders and Isolation Forests. These models analyze system behavior without predefined labels, identifying deviations from normal operations that may indicate a security breach. For instance, an autoencoder trained on legitimate API calls can flag anomalous requests that deviate from established patterns, potentially signaling an attempted intrusion.

To maximize effectiveness, our threat detection system operates at multiple levels of the AI supply chain. During the data ingestion phase, ML models scrutinize incoming datasets for signs of poisoning or corruption. In the model training phase, they monitor gradient updates and parameter changes for unusual activity that might indicate backdoor attacks. Finally, during deployment and inference, the system analyzes input-output pairs to detect adversarial examples designed to deceive the AI. By integrating these ML components with

blockchain's immutable ledger, our framework provides a comprehensive defense mechanism that adapts to evolving threats while maintaining a verifiable record of all activities.

3.3 *Integration Architecture*

The seamless integration of blockchain and ML components is critical to the success of our framework. Our architecture consists of three primary layers: the Data Layer, the Blockchain Layer, and the ML Layer, each playing a distinct role in securing the AI supply chain.

The Data Layer serves as the foundation, storing encrypted datasets, model parameters, and other AI assets. All data entries are hashed and linked to the blockchain, ensuring their integrity and provenance. Before any data is processed, it undergoes validation checks to confirm its authenticity and compliance with predefined standards.

The Blockchain Layer acts as the system's backbone, managing access control, logging transactions, and enforcing smart contracts. This layer ensures that all modifications to the AI supply chain are recorded transparently and immutably. For example, when a new model version is deployed, the blockchain logs the update along with metadata such as the developer's identity, timestamp, and hash of the model file. This enables stakeholders to verify the model's lineage and detect unauthorized changes.

The ML Layer provides real-time threat detection by analyzing data flows and system behavior across the supply chain. This layer interfaces with both the Data and Blockchain Layers, using insights from the blockchain to enhance its detection capabilities. For instance, if the blockchain records an unusual spike in model updates from a particular user, the ML system can investigate further to determine whether this activity is legitimate or malicious. Conversely, if the ML layer detects an anomaly, it can trigger a smart contract to quarantine affected assets until the issue is resolved.

By orchestrating these layers into a cohesive system, our framework delivers end-to-end security for AI supply chains. The blockchain ensures transparency and immutability, while ML enables adaptive and real-time threat detection. Together, they create a robust defense mechanism that addresses the unique challenges of securing AI ecosystems.

4. Experimental Evaluation

To validate the effectiveness of our proposed framework, we conducted a series of experiments using real-world datasets and simulated attack scenarios. The evaluation focused on three key metrics: detection accuracy, false positive rate, and response time. Our test environment comprised a permissioned blockchain network built on Hyperledger Fabric and multiple ML models for threat detection, including LSTM networks for sequential anomaly detection and Isolation Forests for outlier identification.

We utilized two primary datasets for our experiments: the CICIDS2017 dataset for intrusion detection and the TrojAI dataset for model poisoning attacks. The CICIDS2017 dataset provided a comprehensive collection of network traffic logs, including both benign and malicious activities, which allowed us to train and test our ML models on realistic attack scenarios. The TrojAI dataset, on the other hand, contained examples of Trojan attacks on machine learning models, enabling us to evaluate our framework's ability to detect and mitigate model poisoning.

Our results demonstrated significant improvements over traditional security mechanisms. The integrated blockchain-ML framework achieved a detection accuracy of 95.2%, compared to 78.5% for conventional methods. This improvement can be attributed to the synergy between blockchain's immutable record-keeping and ML's adaptive threat detection capabilities. For instance, the blockchain layer provided verifiable context for each data point, allowing the ML models to make more informed decisions about potential threats. Additionally, the framework maintained a false positive rate of just 3.1%, significantly lower than the 12.4% observed in baseline systems. This reduction is critical for minimizing unnecessary disruptions in operational environments.

Response time was another area where our framework excelled. The average time to detect and respond to threats was 0.8 seconds, compared to 2.5 seconds for traditional systems. This speed is essential for mitigating attacks in real time, particularly in high-stakes applications such as autonomous vehicles or healthcare diagnostics. The experiments also revealed that our framework was particularly effective at identifying zero-day attacks, thanks to the unsupervised learning components that could detect anomalies without prior knowledge of specific attack signatures.

In summary, the experimental evaluation confirmed that our integrated approach offers substantial advantages over existing solutions. By combining blockchain's transparency with ML's predictive power, the framework not only enhances security but also maintains operational efficiency, making it a viable solution for securing AI supply chains in practice.

5. Security Analysis

A thorough security analysis is essential to understand the robustness of our framework against various attack vectors. We examined the system's resilience to data tampering, model theft, and adversarial attacks, as well as its ability to maintain confidentiality, integrity, and availability—the core tenets of cybersecurity.

First, the immutable nature of the blockchain ensures that once data or model parameters are recorded, they cannot be altered without detection. This property is particularly valuable for preventing data tampering, as any unauthorized changes would require consensus across the decentralized network, which is computationally infeasible in a well-designed blockchain system. Furthermore, the use of cryptographic hashing guarantees that even minor modifications to the data would result in a completely different hash value, making tampering immediately apparent.

Second, our framework's ML-based threat detection component provides an additional layer of defense against sophisticated attacks. For example, adversarial attacks that attempt to deceive AI models by introducing subtle perturbations to input data can be detected through anomaly detection algorithms. These algorithms analyze input patterns and flag deviations from expected behavior, enabling the system to reject malicious inputs before they can affect the model's output. Similarly, model theft attempts are mitigated through the blockchain's access control mechanisms, which log all access requests and can trigger alerts if unauthorized attempts are detected [17].

Third, the decentralized architecture of our framework enhances its resilience against single points of failure. Unlike centralized systems, where a breach in one component can compromise the entire system, our

distributed approach ensures that even if one node is attacked, the rest of the network remains secure. This is particularly important for maintaining availability, as it prevents attackers from disrupting the entire supply chain by targeting a single vulnerability.

However, no system is entirely impervious to attacks, and our framework is no exception. One potential limitation is the computational overhead associated with blockchain operations, which could impact performance in high-throughput environments. Additionally, while the ML components are highly effective at detecting known and unknown threats, they are not infallible and may occasionally produce false negatives. Future iterations of the framework could address these challenges by optimizing the blockchain consensus mechanism and incorporating more advanced ML techniques, such as federated learning for decentralized threat detection.

Overall, the security analysis confirms that our framework provides a robust defense against the most pressing threats to AI supply chains. By leveraging the complementary strengths of blockchain and ML, it offers a comprehensive solution that is both resilient and adaptable to evolving security challenges.

6. Challenges and Future Work

While our framework demonstrates significant promise in securing AI supply chains, several challenges remain to be addressed to ensure its widespread adoption and long-term effectiveness. These challenges span technical, regulatory, and operational domains, each requiring careful consideration and innovative solutions.

One of the primary technical challenges is the computational and storage overhead associated with blockchain technology. The process of recording every transaction on an immutable ledger, while essential for security, can introduce latency and increase resource consumption. This is particularly problematic for large-scale AI systems that process vast amounts of data in real time. Future research could explore lightweight blockchain protocols or off-chain storage solutions to mitigate these issues without compromising security. Another technical hurdle is the scalability of ML-based threat detection. As AI supply chains grow in complexity, the volume of data to be analyzed increases exponentially, potentially overwhelming the ML models. Techniques such as edge computing and distributed ML could help distribute the computational load and improve efficiency.

On the regulatory front, the lack of standardized security and compliance frameworks for AI supply chains poses a significant challenge. Different industries and jurisdictions may have varying requirements for data privacy, model transparency, and accountability. Our framework would benefit from alignment with emerging standards such as the NIST AI Risk Management Framework and ISO/IEC 27001 for information security. Future work could involve collaborating with policymakers to develop guidelines that ensure our solution meets global regulatory expectations while remaining flexible enough to adapt to evolving threats [18] especially in security as well as in anti-money laundering [20].

Operational challenges include the integration of our framework with existing AI pipelines. Many organizations have already invested heavily in proprietary AI infrastructure, and retrofitting these systems to accommodate blockchain and ML-based security could be complex and costly. To address this, we plan to develop modular and interoperable components that can be seamlessly integrated into diverse environments.

Additionally, user education and training will be critical to ensure that stakeholders understand how to leverage the framework effectively.

Looking ahead, several exciting directions for future research emerge. One promising avenue is the incorporation of quantum-resistant cryptography to safeguard against future threats posed by quantum computing. Another is the exploration of federated learning techniques to enable secure, decentralized model training without centralized data aggregation. Finally, automated incident response mechanisms could be enhanced to not only detect threats but also autonomously mitigate them, further reducing the burden on human operators [19].

In conclusion, while our framework represents a significant step forward in securing AI supply chains, ongoing innovation and collaboration across disciplines will be essential to address the challenges outlined above. By continuing to refine and expand our approach, we can help build a future where AI systems are both powerful and secure, enabling their safe and responsible deployment across industries.

7. Conclusion

The security of AI supply chains is a critical concern as these systems become increasingly integral to modern infrastructure. This paper has presented a comprehensive framework that combines blockchain technology and machine learning-based threat detection to address the unique vulnerabilities inherent in AI ecosystems. By leveraging blockchain's immutable ledger for transparent and tamper-proof record-keeping and ML's adaptive capabilities for real-time anomaly detection, our solution provides end-to-end security across the entire AI lifecycle—from data provenance to model deployment.

Our experimental results underscore the effectiveness of this integrated approach, demonstrating a 15-20% improvement in attack detection accuracy compared to traditional methods. The framework's ability to reduce false positives and accelerate response times further highlights its practical utility in real-world applications. Moreover, the security analysis confirms its resilience against a wide range of threats, including data tampering, model theft, and adversarial attacks [21].

However, as discussed, challenges such as computational overhead, regulatory compliance, and system integration remain. These issues present opportunities for future research, particularly in the areas of quantum-resistant cryptography, federated learning, and automated incident response [22]. Addressing these challenges will be essential to ensuring the framework's scalability and adaptability in the face of evolving threats.

In closing, this research contributes to the broader goal of building trustworthy and secure AI systems. By providing a robust, transparent, and adaptive solution for safeguarding AI supply chains, our framework not only mitigates current risks but also lays the foundation for future advancements in AI security. As AI continues to transform industries, the need for such comprehensive security measures will only grow, making this work a vital step toward a safer and more resilient digital future.

References

1. R. Kumar, P. Kumar, M. Aloqaily, and A. Aljuhani, "Deep-learningbased blockchain for secure zero touch networks," *IEEE Commun. Mag.*, vol. 61, no. 2, pp. 96–102, Feb. 2023. [Online]. Available: https://doi.org/10.1109/MCOM.001.2200294
2. U. Bodkhe, "Blockchain for industry 4.0: A comprehensive review," IEEE Access, vol. 8, pp. 79764–79800, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2988579
3. Zhang, L. Zhu, and C. Xu, "BPAF: Blockchain-enabled reliable and privacy-preserving authentication for fog-based IoT devices," IEEE Consum. Electron. Mag., vol. 11, no. 2, pp. 88–96, Mar. 2022. [Online]. Available: https://doi.org/10.1109/MCE.2021.3061808
4. Attkan and V. Ranga, "Cyber-physical security for IoT networks: A comprehensive review on traditional, blockchain and artificial intelligence based key-security," Complex Intell. Syst., vol. 8, no. 4, pp. 3559–3591, 2022.
5. S. Aghapour, M. Kaveh, M. R. Mosavi, and D. Martín, "An ultralightweight mutual authentication scheme for smart grid two-way communications," IEEE Access, vol. 9, pp. 74562–74573, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3080835
6. S. Jiang, X. Zhu, and L. Wang, "An efficient anonymous batch authentication scheme based on HMAC for VANETs," IEEE Trans. Intell. Transp. Syst., vol. 17, no. 8, pp. 2193–2204, Aug. 2016. [Online]. Available: https://doi.org/10.1109/TITS.2016.2517603
7. F. J. Kurfess, "Artificial intelligence," in *Encyclopedia of Physical Science and Technology*, 3rd ed., R. A. Meyers, Ed. New York, NY, USA : Academic, 2003, pp. 609–629.
8. Y. K. Dwivedi, "Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manag.*, vol. 57, Apr. 2021, Art. no. 101994.
9. J. Frizzo-Barker, P. A. Chow-White, P. R. Adams, J. Mentanko, D. Ha, and S. Green, "Blockchain as a disruptive technology for business: A systematic review," *Int. J. Inf. Manage.*, vol. 51, Apr. 2020, Art. no. 102029.
10. Y. Yuan and F.-Y. Wang, "Blockchain and cryptocurrencies: Model, techniques, and applications," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 48, no. 9, pp. 1421–1428, Sep. 2018.
11. S. Nakamoto. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Accessed: Jul. 24, 2023. [Online].
12. J. A. Jaoude and R. George Saade, "Blockchain applications—Usage in different domains," *IEEE Access*, vol. 7, pp. 45360–45381, 2019.
13. A. A. Khan, A. A. Laghari, M. Rashid, H. Li, A. R. Javed, and T. R. Gadekallu, "Artificial intelligence and blockchain technology for secure smart grid and power distribution automation: A state-of-the-art review," *Sustain. Energy Technol. Assessments*, vol. 57, Jun. 2023, Art. no. 103282.
14. S. Kumar, W. M. Lim, U. Sivarajah, and J. Kaur, "Artificial intelligence and blockchain integration in business: Trends from a bibliometric-content analysis," *Inf. Syst. Frontiers*, vol. 25, no. 2, pp. 871–896, 2023.
15. N. Tsolakis, R. Schumacher, M. Dora, and M. Kumar, "Artificial intelligence and blockchain implementation in supply chains: A pathway to sustainability and data monetisation?," *Ann. Oper. Res.*, vol. 327, no. 1, pp. 157–210, Aug. 2023.
16. S. Selvarajan, G. Srivastava, A. O. Khadidos, A. O. Khadidos, M. Baza, A. Alshehri, and J. C.-W. Lin, "An artificial intelligence lightweight blockchain security model for security and privacy in IIoT systems," *J. Cloud Comput.*, vol. 12, no. 1, p. 38, Mar. 2023.
17. E. Bertino, A. Kundu, and Z. Sura, "Data transparency with blockchain and AI ethics," *J. Data Inf. Qual.*, vol. 11, no. 4, pp. 1–8, Dec. 2019.
18. S. Vyas, M. Shabaz, P. Pandit, L. R. Parvathy, and I. Ofori, "Integration of artificial intelligence and blockchain technology in healthcare and agriculture," *J. Food Qual.*, vol. 2022, pp. 1–11, May 2022.
19. R. Kumar, D. Singh, K. Srinivasan, and Y.-C. Hu, "AI-powered blockchain technology for public health: A contemporary review, open challenges, and future research directions," *Healthcare*, vol. 11, no. 1, p. 81, Dec. 2022
20. Rajpoot, M. H., & Raffat, M. W. (2024). The AI-Driven Compliance and Detection in Anti-Money Laundering: Addressing Global Regulatory Challenges and Emerging Threats: AI-Driven AML: Compliance Threat Detection. Journal of Computational Science and Applications (JCSA), ISSN: 3079-0867 (Onilne), 1(2).

21. Ahmad, Z., Obaidullah., Ashraf, M. A., & Tufail, M. (2024). Enhanced Malware Detection Using Grey Wolf Optimization and Deep Belief Neural Networks. International Journal for Electronic Crime Investigation, 8(3).
22. Iqbal, M., Shafiq, M. U., Khan, S., Obaidullah, Alahmari, S., & Ullah, Z. (2024). Enhancing task execution: a dual-layer approach with multi-queue adaptive priority scheduling. PeerJ Computer Science, 10, e2531.