LLM Supply Chain Provenance: A Blockchain-Based Approach

Shridhar Singh and Luke Vorster
University of KwaZulu-Natal, Westville, South Africa

217008024@stu.ukzn.ac.za
VorsterL@ukzn.ac.za

Abstract: The burgeoning size and complexity of Large Language Models (LLMs) introduce significant challenges in ensuring data integrity. The proliferation of "deep fakes" and manipulated information raises concerns about the vulnerability of LLMs to misinformation. Traditional LLM architectures often lack robust mechanisms for tracking the origin and history of training data. This opaqueness can leave LLMs susceptible to manipulation by malicious actors who inject biased or inaccurate data. This research proposes a novel approach integrating Blockchain Technology (BCT) within the LLM data supply chain. With its core principle of a distributed and immutable ledger, BCT offers a compelling solution to address this challenge. By storing the LLM's data supply chain on a blockchain, we establish a verifiable record of data provenance. This allows for tracing the origin of each data point used to train the LLM, fostering greater transparency and trust in the model's outputs. This decentralised approach minimises the risk of single points of failure and manipulation. Additionally, the immutability of blockchain records ensures that the data provenance remains tamper-proof, further enhancing the trustworthiness of the LLM. Our approach leverages three critical features of BCT to strengthen LLM security: 1) Transaction Anonymity: While data provenance is recorded on the blockchain, identities of data contributors can be anonymised, protecting their privacy while ensuring data integrity. 2) Decentralised Repository: Enhances the system's resilience against potential attacks by distributing the data provenance record across the blockchain network. 3) Block Validation: Rigorous consensus mechanisms ensure the validity of each data point added to the LLM's data supply chain - minimising the risk of incorporating inaccurate or manipulated data into the training process. Using the experimental approach, initial evaluations using simulated LLM training data on a blockchain platform demonstrate the feasibility and effectiveness of the proposed approach in enhancing data integrity. This approach has far-reaching implications for ensuring the trustworthiness of LLMs in various applications.

Keywords: Blockchain-based data provenance, LLM security, Generative models

1. Introduction

The progression of Generative AI (GenAI) technology, particularly Large-Language Models (LLMs), has revolutionised many fields. From machine translation and content creation to chatbot development, LLMs have changed how we use AI and continue to shape the future of AI-powered tools. As such, the security and trustworthiness of LLMs are paramount. Untrustworthy LLM provenance, meaning a lack of transparency about the data used to train the model, can introduce biases, facilitate manipulation, spread propaganda, and create security vulnerabilities (Singh, 2024; Wang *et al.*, 2024). For instance, an LLM trained on biased data could perpetuate societal stereotypes in its outputs. Traditional methods for LLM securing supply chains may not be sufficient for the complexities of LLM development, highlighting the need for a more robust approach (Luo, Luo and Vasilakos, 2023; Wang *et al.*, 2024; Zuo *et al.*, 2024).

Blockchain technology (BCT), designed to provide integrity and safety within the fintech sector, provides a compelling case for robust security integration within LLMs. With its core principles of a distributed ledger, immutability, and secure consensus mechanisms, BCT offers a promising solution for enhancing LLM supply chain provenance, promoting decentralisation and immutability (Treiblmaier, 2018; Himeur *et al.*, 2022). By leveraging blockchain, we can create a transparent and tamper-proof record of every step in the LLM lifecycle, from data collection and model training to end-user interactions and fine-tuning (Luo, Luo and Vasilakos, 2023; Zuo *et al.*, 2024). We hypothesise/argue that this increased transparency can significantly improve trust and confidence in the integrity of LLMs.

Building upon our previous work on securing LLMs with zero-knowledge proofs, this research investigates the potential of BCT to enhance LLM supply chain provenance (Singh, 2024). Our research focuses on designing and conducting a series of simulations assessing a blockchain-based LLM's core functionalities. By leveraging existing LLM architectures and blockchain development techniques, we can create a simulated environment to explore enhanced security, scalability, and trustworthiness through transparency.

Through these simulations, we aim to address the following research questions:

RQ1: How can BCT be implemented to create a secure and verifiable record of data provenance throughout the entire LLM data supply chain?

RQ2: What specific data points within the LLM supply chain are most critical to track and verify using blockchain?

RQ3: How can BCT mitigate the risks of bias, manipulation, and security vulnerabilities introduced by untrustworthy LLM provenance?

RQ4: How can blockchain-based provenance systems be designed to ensure scalability and interoperability for large-scale LLM deployment?

RQ5: How can blockchain-based provenance systems be designed to promote transparency and trust among stakeholders involved in the LLM development process?

Our work aims to ensure verifiable and secure data provenance throughout the LLM lifecycle, drawing inspiration from the architectural solutions proposed by Zuo *et al*. (2023). By analysing the results of these simulations, we aim to contribute valuable insights into the feasibility and effectiveness of blockchain-based LLMs. This knowledge can inform future research and development efforts to create secure, trustworthy, and transparent LLMs empowered by blockchain technology.

This serves as a cornerstone paper outlining the trajectory for our research and future research endeavours on blockchain-based LLM security. As such, concerns and questions outlined in this paper may only partially be answered or addressed within the paper itself and will be the focus of our future work on the topic. It is essential to highlight the broad scope of the research at the outset to plan a road map of objectives we set out to achieve.

2. Literature Review

This literature review focuses on the emerging concept of LLM supply chain provenance and explores the potential of blockchain technology in addressing its challenges. Focusing on research published within the last eight years (2016-2024), this review aims to identify key themes in LLM supply chain research, analyse the potential of blockchain technology, and identify gaps in the research.

2.1 Trust in LLM Supply Chains

As LLMs become more prevalent, concerns regarding data bias, security vulnerabilities, and the interpretability of their outputs are paramount (Balayn *et al.*, 2024). Therefore, the need arises to create more robust LLM supply chains to facilitate this data transfer to and from the LLM on large-scale applications, such as Large Multimodal Model (LMM) applications. Defined as a simple three-phase circuit of events comprising LLM Infrastructure, Model Lifecycle, and Application Ecosystem, the LLM supply chain runs deep into the LLM training and data transfer processes where it monitors data flow from the infrastructure development level to the model lifecycle and finally to the end application ecosystem (Wang *et al.*, 2024).

This supply chain is responsible for the trust processes of the LLM, both human and non-human entities. As multiple organisations or individuals are likely to be involved in the LLM development, these trustor and trustee relationships shape the supply chain as producers, consumers, or indirect stakeholders (Balayn *et al.*, 2024). Should these supply chain actors not adhere to strict protocols regarding data cleaning and ethical and transparent operational processes, they could taint the LLM to adhere to biases which would affect the user of the LLM-powered application (Wang *et al.*, 2024).

Establishing trust in these systems, therefore, relies on verifiable data provenance and secure training processes (Balayn *et al.*, 2024; Wang *et al.*, 2024). Trustworthy AI requires explainability, fairness, and robustness of trust-based LLM solutions (Singh, 2024). Therefore, understanding the origin and characteristics of training data allows for identifying and mitigating potential biases, promoting fairer AI development. Secure training processes and data transfer are crucial to prevent malicious actors from manipulating data and compromising the integrity of AI models and their downstream applications (Luo, Luo and Vasilakos, 2023; Wang *et al.*, 2024).

2.2 Blockchain-Based Solutions for LLM Supply Chain

Invented in 2008 by Satoshi Nakamoto, Blockchain is a distributed, immutable ledger built for tracking transactions and providing an un-tamper-able, verifiable provenance record (Guo and Yu, 2022). Primarily used in fintech, this technology tracks the transactions of digital currency and cryptocurrency between entities and records these transactions in a block. Blocks are then verified by other users, called miners, through consensus mechanisms – a voting method where nodes within the network work to verify the transactions within the block (Himeur *et al.*, 2022). If transactions are valid, the block is added to the Blockchain and will be uneditable. Since the addition of new blocks to the chain relies on the hash values of the immediate previous block, any attempts to alter the chain will cause ripple effects breaking the transactions to follow, making it easy to identify where the tampering occurred (Guo and Yu, 2022; Himeur *et al.*, 2022).

BCT, therefore, presents a promising approach for establishing secure and transparent supply chains (Apte and Petrovsky, 2016; Treiblmaier, 2018) and can be a vital tool to secure LLM supply chains (Luo, Luo and Vasilakos, 2023; Zuo *et al.*, 2024). Luo, Luo and Vasilakos (2023) propose a framework called BC4LLM that leverages blockchain to ensure the security of the entire LLM training process, including creating a reliable learning corpus by verifying data ownership ensuring data security, and facilitating secure training processes. However, two drawbacks of this solution are reduced scalability and high energy consumption.

Similarly, Fan *et al.* (2023) introduced a federated learning framework called FATE-LLM, offering a promising approach for addressing data privacy and security concerns within the LLM supply chain. Federated learning allows for training LLMs on distributed datasets without requiring data to be transferred to a central location. This approach mitigates the risk of data breaches and empowers data owners to maintain control over their data. However, federated learning also presents challenges, including communication overhead between training clients and the central server.

2.3 Trusting and Scaling Blockchain-Based Approaches

Blockchain-based approaches to cloud and other distributed computing platforms see a lack of consumer confidence due to lack of transparency and loss of control over data (Gong and Navimipour, 2022). Further potential limiting factors to the adoption of Blockchain-based approaches include lack of degraded networking control, susceptibility to majority attacks, and irreversible commits to the blockchain (Zhu *et al.*, 2019). Gong and Navimipour (2022) highlights research suggesting possible solutions to these problems such as Controllable Blockchain Data Management which implements a Trust Authority node to defend against majority attacks (Zhu *et al.*, 2019), and scalability and reliability of blockchains assessed by Rimba *et al.* (2020) found that their cost model strengthened their outcome reliability and helped justify their costs.

Further, the Blockchain-enabled IoT approach explored by Singh, Rathore and Park (2020) explore AI-based strategies to minimise the computational overhead required for blockchain mining. The approach addresses the blockchain architecture and investigates the use of path discovery, data association, prediction, aggregation, and classification to improve the performance of blockchain miners, thus decreasing the mining cost and computational overhead required. Additionally, this integration can be effective in increasing the scalability of the solution, essentially fending off 51% majority attacks in the process (Singh, Rathore and Park, 2020). To further address the issue of scalability, the researchers again approach the problem using federated learning to distribute the sheer capability of the system as opposed to opting for a centralised approach to data storage.

The literature addresses approaches to facilitate LLM secure training processes, mitigate the risk of data breaches, and highlight the potential for scalability challenges of Blockchain-based LLMs. However, further research is required to understand these frameworks' practical applications and limitations regarding scalability, adoption, and field testing. Additionally, research thus far has yet to address the challenge of how untrustworthy actors can be identified and dealt with from within the Blockchain itself. The verification and validation of sources rely on the source creator's trust, and the accuracy of the information cannot be validated before it is used in the training process. No framework highlighted the explicit dataset an LLM utilised to generate its response. These are all issues we aim to address in this research.

3. Methodology

This chapter outlines the methodology employed in our research to explore Blockchain-based LLMs' potential. As our focus lies in understanding the practical functionalities of this technology, we adopted an empirical approach centred on simulation and evaluation.

3.1 Research Objectives

Design and conduct simulations that evaluate the functionalities of a Blockchain-based LLM to:

- Investigate the feasibility of utilising BCT to create a verifiable and secure record of data provenance throughout the LLM lifecycle.
- Explore how Blockchain can mitigate the risks associated with untrustworthy LLM provenance, such as bias and manipulation.
- Analyse the potential of blockchain-based LLMs to enhance security, scalability, and trust within the LLM development process.

3.2 Simulation Design

3.2.1 Core components

Our simulations centred on two core components:

- **LLM Architecture:** Drawing inspiration from open-source libraries to influence our design, we achieved our objective of simulating core LLM functionalities, leveraging a wide variety of open-source LLMs and allowing us to choose custom embeddings and vector databases (obtained from HuggingFace).
- **Blockchain Development:** We developed a custom-simulated blockchain architecture to monitor the various components of the supply chain using established blockchain development techniques.

Simulation Scenarios:

Simulations were designed to assess the provenance tracking ability of blockchain-based LLMs and the database detection and mitigation strategies that they offer. Simulations scenarios were grouped into three categories:

- Simulation 1: **Verifiable data provenance**: This scenario simulated data flow throughout the LLM supply chain. Relevant data points such as model name, embeddings, vector database utilised, etc., were logged as entries on the blockchain. This provides a verifiable chain of events leading to response generation.
- Simulation 2: **Bias mitigation**: This scenario explored the bias mitigation techniques implemented within the blockchain-based LLM supply chain. Analysis of sources and data categorisation are critical components of this simulation.
- Simulation 3: **Data manipulation and Security Analysis**: This scenario focused on simulating security breaches and unauthorised access attempts. We evaluated the effectiveness of blockchain's security features in preventing unauthorised data manipulation and protecting the integrity of the LLM.

3.3 Data Collection and Curation

Since this research focuses on assessing the viability of Blockchain-based LLMs rather than on the overall training and fine-tuning process, we have chosen to opt for a custom dataset comprising 200 academic research papers. This allows us to effectively construct simulations of adding and modifying source data and evaluate Blockchain security's effectiveness. Thus, this requires more than a few static datasets and must involve a multitude of text-based data with verifiable sources, even if the total size of the data is smaller.

Data on the simulation design of Blockchain-based LLMs were collected by consolidating information from the literature review with our series of trial-and-error designs. Each design was evaluated and contributed towards the final simulation on which we base our findings.

4. Findings

This chapter explores the implementation of BCT to create a secure and verifiable record of data provenance throughout the entire LLM lifecycle.

Simulation 1: Verifiable data provenance and crucial supply chain data points

Blockchain design and security architecture (RQ1)

Our blockchain architecture focused on modularity and utilised three chain types: Data Vendor chain, Data Categorisation chains, and Master chain. Each block comprises the following standard components: block hash, block header, previous block hash, Merkle root, nonce value, and timestamp. Our design leverages the permissioned Blockchain architecture, requiring all users to provide credentials before accessing any functionality. User credentials and every other data point monitored by the blockchain utilise the secure one-way cryptographic hashing algorithm, SHA256, to ensure anonymity whilst preserving the data in a tamper-proof format. Any alterations to these data points will generate a new hash, prohibiting changes once the data point is logged and passed to the blockchain.

Leveraging multithreading to represent worker nodes on our simulated distributed network, we stipulate that nodes must evaluate the chain's current state before any new data can be written to the blockchain. This ensures uniformity of the chain state and protects against chain corruption. Data points representing interactions with the LLM are then stored on a block; these interactions vary depending on the chain – Master, Data Vendor, or Data Categorisation. Unique data points for each chain are listed in Table 1 below. These blocks are mined by

the network nodes utilising the Proof-of-Work consensus mechanism. Once a block is successfully mined, it is added to its respective chain, where it lives in an un-tamper-able state, as any alterations to this entry will derail future transactions.

Blockchain and the LLM supply chain (RQ1 & RQ2)

Our design builds on the data collected from the literature and specifies unique data points that represent the individual LLM-specific aspects. Table 1 highlights the data points for the three chains used in this version of our simulation.

Table 1: Shows the different chain types and which data points each chain monitors

UNIQUE DATA POINTS	
MASTER CHAIN	Model
	Embeddings
	Vector database
	Vendor
DATA CHAIN	Category
VENDOR CHAIN	Vendor
DATA CATEGORISATION CHAIN	Category

The Blockchain tracks LLM source data upload through the Data Vendor and relevant Data Categorisation chains. The Data Vendor chain tracks the entity that supplied the data and the specific Data Categorisation chain logs where the data is stored. When more data are added to any chain, the same mining principles apply, and a previous state cannot be reverted to or altered without affecting succeeding blocks. Only once all blocks related to data supply are successfully mined, the data get added to the vector database for the LLM to consult when answering user prompts.

When a user interacts with the LLM, the chat interaction gets logged on the Master chain. Data such as model used, embeddings, chat history, vector database, temperature, data vendor, source data category, and user wallet ID are captured within a block.

Simulation 2: Bias Mitigation (RQ3)

This simulation assesses Blockchain's ability to identify and prevent potential biases that supply chain actors have on LLM outcomes. This research defines bias as any overreliance or undue influence on a particular outcome. To achieve this, we set out the following preconditions we believe influence potentially biased outcomes:

1. LLM's overreliance on specific data sources
2. Source database lacks sufficient data on a topic
3. Source database does not include data from enough vendors on topics
4. References provided by LLM could not be validated
5. High temperature settings of the LLM

This simulation leverages the blockchain architecture pinning it against the stock LLM to determine whether the Vendor and Data Categorisation chains benefit the LLM in bias mitigation. This simulation was conducted with these chains not active, *Case 1*, and active, *Case 2*, to determine a before and after result. Prompts and datasets were kept constant, and temperature settings were incrementally increased by 0.1 degrees, starting at a temperature of 0.1 until a peak of 1.0.

Table 2: Shows the average of results from 200 total prompts

	Responses Containing Bias /Per 10 Prompts	Unverified Sources /Per 10 Prompts	Repeat Sources /Per 10 Prompts	Total Sources Consulted /Per 10 Prompts
Case 1 – Chains Not Active	3.46	1.1	2.36	4.16
Case 2 – Chains Active	1.42	0.2	1.22	4.86

As shown in table 2 above, from the 200 prompts inputted, Case 1 responses were particularly susceptible to influential bias, with an average of 3.46 responses containing bias, per 10 prompts. The most significant contributor to this metric was referencing sources that could not be verified at a rate of 1.1 unverified sources for every ten responses. The LLM also displayed an overreliance on a handful of sources, listing an average of 2.36 repeat sources out of 4.16 total sources consulted, per 10 responses.

Case 2 generated contrasting results as we observed the influential bias of responses decrease to an average of 1.42 responses containing bias, per 10 prompts. Unverified sources were reduced to 0.2, per 10 prompts, and repeat sources were, on average, 1.22 out of 4.86 total sources consulted, per 10 responses.

Simulation 3: Data Manipulation and Security Analysis (RQ3 & RQ1)

Table 3 details desired and actual outcomes of the security analysis by stipulating the alterations made to the blockchain and their potential implications. The ideal outcome of this experiment is to have a score of zero, as any higher would indicate a potential vulnerability.

Table 3: Shows outcomes of security analysis testing conducted

Types of alterations	Proposed effect on the Blockchain (desired outcome)	How it affected the Blockchain (actual outcome)	Implication weight	Actual Weight
Block Hash	Disrupt the flow of data between user-LLM into the blockchain and alter new and existing transactions	Every instance of tampering related to the block hash manipulation was successfully detected and stopped	5	0
Nonce	Attempt to alter a new transaction not yet mined to represent and change properties not related to user-LLM interaction	No alteration of transactions was successfully added to the blockchain	5	0
Source Dataset	Infect the source information and introduce bias/misinformation into the LLM's response computation	Any changes to source data were detected before the LLM execution began	5	0
Timestamp	Lesser attempt but an indication that there were efforts to sabotage the blockchain	No alteration of timestamp was recorded on the blockchain	2	0
New (malicious) data source	Attempt to introduce a malicious actor to supply information	The system allowed the creation of a new actor	3	3
51% attack	Attempt to overwhelm the blockchain by staking a claim that majority of the nodes agree with the altered state of transactions as opposed to the original	This typical attack was neutralised by the simulation keeping the blockchain state in memory	4	0

From this table, we can detect one source of potential vulnerability with the blockchain pertaining to source creation.

5. Discussion

This chapter delves into the implications of the results presented, exploring the potential of BCT for establishing a secure and verifiable record of data provenance throughout the LLM lifecycle.

These findings hold significant value for the development of secure and trustworthy LLMs. Establishing a verifiable record of data provenance allows stakeholders to gain insights into the data used to train LLMs,

fostering transparency and accountability within the LLM development process. By mitigating risks like manipulation and bias, BCT can contribute to producing more reliable and unbiased LLM outputs.

5.1 Evaluation of Findings

Research Question 1: Implementing BCT for Secure and Verifiable Provenance Tracking

To assess the verifiable record of data provenance, we adopted an object-oriented software approach to our blockchain design. Leveraging the concept of abstraction, we designed the blockchain to accommodate various individual components that symbolise the flow of data. This flow of data is monitored on a master chain, whilst smaller individual chains handle the responsibility of security and integrity of data sources and the categorisation of data for easily monitoring any amendments to the data and source catalogues of the LLM.

Ensuring a verifiable record of transactions requires a clear understanding of where each transaction is going, meaning what task is carried out by executing that transaction. We have broken down the transaction process into three streams of data flow: an input stream of data, a verification stream, and a chat stream.

The data input stream begins with vendor validation, as identified within the literature review; data streams may not be from the same source. Data quality is crucial in ensuring accurate and reliable LLM computations, and vendor validation is essential to ensure accountability for the data supplied to the LLM.

Research Question 2: Tracking and verifying critical supply chain data

Our simulation design evolved as we conducted this research, and more demand was placed on the blockchain. We initially identified the blockchain components based on the findings within the literature and concluded that a more robust solution was required. We created a verifiable architecture for our final implementation using three blockchains: Vendor, Data Categorisation, and Master chain. The standard and unique data points identified for use within these chains provide crucial provenance tracking capabilities that allowed for the use of this design in answering other questions highlighted in this research.

Although further research is required, this approach to architectural design provides a solid foundation for scalable solutions as the demands of the blockchain increase and applications require domain-specific knowledge.

Research Question 3: Mitigating Risks with BCT

Simulation 2 allowed us to focus on the ability of the blockchain architecture to provide meaningful contributions to bias mitigation within the LLM response generation process. As per the definition of bias described above, bias in the context of LLMs can stem from overreliance on specific data points and sources to answer a particular type of prompt. As such, skewed or unjust data providers can input false data if not verified.

By ensuring the LLM consults more than one source of information before generating its response, we can limit the potential for bias to creep into the LLM's output should the consulted data be skewed in any manner. This is especially useful in scenarios of potential vulnerability where the data vendor is valid, but the source data is only partially bias-free.

Simulation 2 shows that BCT's architecture provides a solid foundation for data vendors to provide legitimate and accurate data and reduces the challenges associated with interdependencies of, and interactions with the various supply chain actors (Wang *et al.*, 2024). Thus, this aids in prohibiting unwanted bias/influence on the LLMs response generation and provides a more significant deal of trust that the data and output will be trustworthy.

Simulation 3 demonstrated the potential of BCT to mitigate security risk manipulation. The immutability of the blockchain ledger prevented most attacks relating to data manipulation however further investigation is required on how to best prevent malicious data vendors from entering the network.

Research Question 4: Scalability and Interoperability for large-scale LLM development

Although not explored experimentally in this work, as BCT platforms already host fair degrees of relative scalability, were this limit reached in the context of considerably large LLMs, we can see a strong argument for the innovation of hybrid blockchain technology - utilising low-computation-cost blockchains to supplement a larger blockchain where expansion of capacity would be required.

Research Question 5: Promote transparency among stakeholders

Although not explored experimentally, our current research suggests the use of transparency-chains to monitor and ensure that each occurred transaction abides by the regulations stipulated by an organisation. Blockchain-based LLM solutions can therefore be tailored to meet organisational needs.

5.2 Limitations of This Study

Blockchain assumes that data added to the chain is correct and final. Thus, Blockchain does not allow any amendments to existing data on the chain. Therefore, this study did not take the action of updating or removing chain-data into consideration. Further, the current representation depicts that we have a solution for an immutable supply chain – essentially a snapshot of the chain, as the chain must alert us when a change occurs. However, the supply chain is not stagnant – it grows and changes with the addition of new data sources. Future work will consider expanding this chain to include elements, such as web supplementation, where every time the LLM consults a new source, it creates a new chain.

5.3 Recommendations and Future Work

This research lays the groundwork for further exploration of BCT in LLM provenance tracking. Future directions include:

- Scalable Permissioned Blockchain Architecture: Investigating and implementing BFT protocols to enable a robust and scalable network for real-world LLM training scenarios.
- Enhanced Data Validation Techniques: Integrating machine learning models for data quality assessment and anomaly detection within the validator script.
- Integration with LLM Development Tools: Developing APIs to facilitate seamless integration of the BCT system with existing LLM development workflows.
- Smart Contract Implementation: Exploring the development of smart contracts to automate key functionalities like data validation, model lineage recording, and access control on the blockchain.

These future research directions promise a more secure, transparent, and trustworthy LLM development ecosystem. By establishing a verifiable record of data provenance, BCT empowers stakeholders to assess potential biases, identify manipulation attempts, and ultimately foster trust in the outputs generated by LLMs. This, in turn, can pave the way for the responsible and ethical development and deployment of LLMs, unlocking their full potential to benefit society.

6. Conclusion

As GenAI advancements become increasingly complex, the necessity for a robust foundation in cybersecurity becomes an ever-growing challenge (Singh, 2024). LLMs are the cornerstone of many GenAI applications, acting as the engines that translate instructions from human users into tangible outputs. Therefore, securing the LLM supply chain – the entire process from data collection to model deployment – is paramount for ensuring GenAI's trustworthiness and ethical use.

Previous research has explored promising avenues for LLM security, with zero-knowledge proofs demonstrating potential in certain aspects. This work builds upon this foundation by investigating the potential of blockchain technology to enhance LLM supply chain security further. Our core hypothesis centred on the idea that blockchain's core principles – immutability, secure record-keeping, and verifiable transactions – could solve the challenge of untrustworthy LLM provenance.

The results of these simulations offer valuable insights into the feasibility and effectiveness of blockchain-based LLMs. They demonstrate the potential for blockchain technology to create a secure and verifiable record of data provenance throughout the LLM lifecycle, addressing the core research questions outlined at the outset. This research paves the way for architectural solutions integrating blockchain to ensure secure and verifiable data provenance within LLM supply chains. Building upon the groundwork laid by Luo, Luo and Vasilakos (2023) and Zuo *et al.* (2024), future research can delve deeper into specific architectural designs and explore their practical implementation. Ultimately, our work contributes to creating secure, trustworthy, and transparent LLMs empowered by blockchain technology.

Acknowledgement

The authors would like to acknowledge the financial contribution made by Mr Luke Vorster from the School of Mathematics, Statistics, and Computer Science at The University of KwaZulu-Natal, South Africa, for funding the

publication and presentation of this research at the 4th International Conference on Artificial Intelligence Research 2024.

References

Apte, S. and Petrovsky, N. (2016) 'Will blockchain technology revolutionise excipient supply chain management?', *Journal of Excipients and Food Chemicals*, 7(3), pp. 76–78.

Balayn, A., Yurrita, M., Rancourt, F., Casati, F. and Gadiraju, U. (2024) 'An Empirical Exploration of Trust Dynamics in LLM Supply Chains'.

Fan, T., Kang, Y., Ma, G., Chen, W., Wei, W., Fan, L. and Yang, Q. (2023) 'FATE-LLM: A Industrial Grade Federated Learning Framework for Large Language Models'. arXiv.

Gong, J. and Navimipour, N.J. (2022) 'An in-depth and systematic literature review on the blockchain-based approaches for cloud computing', *Cluster Computing*, 25(1), pp. 383–400.

Guo, H. and Yu, X. (2022) 'A survey on blockchain technology and its security', *Blockchain: Research and Applications*, 3(2), p. 100067.

Himeur, Y., Sayed, A., Alsalemi, A., Bensaali, F., Amira, A., Varlamis, I., Eirinaki, M., Sardianos, C. and Dimitrakopoulos, G. (2022) 'Blockchain-based recommender systems: Applications, challenges and future opportunities', *Computer Science Review*, 43, p. 100439.

Luo, H., Luo, J. and Vasilakos, A.V. (2023) 'BC4LLM: Trusted Artificial Intelligence When Blockchain Meets Large Language Models'.

Rimba, P., Tran, A.B., Weber, I., Staples, M., Ponomarev, A. and Xu, X. (2020) 'Quantifying the Cost of Distrust: Comparing Blockchain and Cloud Services for Business Process Execution', *Information Systems Frontiers*, 22(2), pp. 489–507.

Singh, S. (2024) 'Enhancing Privacy and Security in Large-Language Models: A Zero-Knowledge Proof Approach', *International Conference on Cyber Warfare and Security*, 19(1), pp. 574–582.

Singh, S.K., Rathore, S. and Park, J.H. (2020) 'BlockIoTIntelligence: A Blockchain-enabled Intelligent IoT Architecture with Artificial Intelligence', *Future Generation Computer Systems*, 110, pp. 721–743.

Treiblmaier, H. (2018) 'The impact of the blockchain on the supply chain: a theory-based research framework and a call for action', *Supply Chain Management: An International Journal*, 23(6), pp. 545–559. Available at: https://doi.org/10.1108/SCM-01-2018-0029.

Wang, S., Zhao, Y., Hou, X. and Wang, H. (2024) 'Large Language Model Supply Chain: A Research Agenda'.

Zhu, L., Wu, Y., Gai, K. and Choo, K.-K.R. (2019) 'Controllable and trustworthy blockchain-based cloud data management', *Future Generation Computer Systems*, 91, pp. 527–535.

Zuo, X., Wang, M., Zhu, T., Zhang, L., Ye, D., Yu, S. and Zhou, W. (2024) 'Federated TrustChain: Blockchain-Enhanced LLM Training and Unlearning'.