Supporting Trustworthy AI with Enterprise Blockchain

Kimia Soroush
Graduate Research Assistant at Shiraz University
soroush.kimia@cse.shirazu.ac.ir

I. Introduction

In today's fast-paced digital world, Artificial Intelligence (AI) is becoming a part of almost every sector. But as AI grows, so do concerns about whether we can trust these systems. Key issues include how AI handles our data, how transparent the AI models are, and how we ensure they behave reliably and ethically.

To make AI systems trustworthy, we focus on three main areas:

1) **Trustworthy Machine Learning:** Ensuring the AI learns from data in a reliable way.
2) **Explainable AI:** Making AI decisions understandable and transparent.
3) **Trustworthy Data Management:** Keeping data safe, private, and accurate.

Blockchain technology, known for its secure and transparent digital ledger, could be the solution to these trust issues. By combining blockchain and AI, we can better track where data comes from, protect the integrity of AI models, and make AI decision-making more transparent. This combination can help build AI systems that are reliable and trustworthy.

Our research will dive into how blockchain can support trustworthy AI. We will look at how blockchain can improve data handling, machine learning processes, and AI decision-making. This study builds on previous work done at the Centre de Recherche en Informatique de l'Université de Paris 1 Panthéon-Sorbonne and will explore new ways to implement blockchain-enhanced AI systems.

II. Research Questions

1) **Trustworthy Data Management:**
   - How can blockchain be utilized to build a trustworthy data preparation pipeline for AI?
   - What blockchain mechanisms ensure the integrity and provenance of AI training data?
2) **Trustworthy Machine Learning:**
   - How can blockchain enhance the trustworthiness of Federated Learning systems?
   - What blockchain-based approaches can mitigate the risks of data poisoning and model tampering?
3) **Trustworthy Inference (Explainable AI):**
   - How can blockchain support the explainability and transparency of AI inference processes?
   - What roles do smart contracts and decentralized oracles play in achieving explainable AI?
4) **Systemic Trust in AI:**
   - How can the various components of Trustworthy AI be integrated into a cohesive, blockchain-supported system?
   - What end-to-end solutions can blockchain provide to ensure the overall trustworthiness of AI ecosystems?

III. Methodology

A. Parameters Considered in the Research

In this research, we focus on three key areas to ensure AI systems are reliable and ethical. First, **Trustworthy Machine Learning** ensures AI models are fair and transparent, preventing against issues like data manipulation. Second, **Trustworthy Inference** makes AI decisions understandable to humans, fostering user trust. Lastly, **Trustworthy Data Management** involves using secure, high-quality data that respects privacy, forming a solid foundation for AI models. By addressing these aspects, we aim to build AI systems that people can trust.

B. Methodology Used in Each Field

To tackle these challenges, we employ innovative methods across three areas. For **Trustworthy Machine Learning**, we use Federated Learning and integrate blockchain to securely train AI models without sharing sensitive data. For **Trustworthy Inference**, blockchain supports explainable AI by making decision processes transparent and traceable. For **Trustworthy Data Management**, blockchain ensures that data used in AI is high-quality and securely handled through robust tracking and auditing. These approaches safeguard data and enhance the robustness and reliability of AI models.

C. Implementation and Evaluation

We will test our methods in real-world applications, particularly in healthcare and finance, where trust and privacy are vital. We'll develop models that securely manage data, offer clear AI decision explanations, and ensure privacy. Using blockchain, we'll enhance data traceability and security, improving AI learning and decision-making. These models will be evaluated in real industrial scenarios to demonstrate their effectiveness. Our goal is to create a robust, ethical approach for trustworthy AI systems across various industries, ensuring they perform well and earn users' trust.

The research will adopt a mixed-methods approach, combining theoretical modeling, system design, and empirical evaluation.

979-8-3503-6104-9/24/$31.00 ©2024 IEEE

1) **Literature Review and State-of-the-Art Analysis:**
   - Conduct a comprehensive review of current literature on Trustworthy AI and blockchain technology.
   - Identify gaps and opportunities for integrating blockchain to enhance AI trustworthiness.
2) **Design and Development:**
   - Develop blockchain-based models and frameworks for data management, machine learning, and inference.
   - Use theoretical insights and design principles to propose solutions addressing identified research questions.
3) **Implementation and Experimentation:**
   - Implement prototype systems demonstrating the feasibility of the proposed models.
   - Conduct experiments using real-world data and industrial use cases to evaluate the performance and reliability of these systems.
4) **Evaluation and Validation:**
   - Assess the developed systems against predefined trustworthiness metrics such as data integrity, privacy, model fairness, and transparency.
   - Validate the effectiveness of blockchain integration in enhancing the trustworthiness of AI systems through rigorous testing and peer review.

IV. Dissemination Strategy

- **Academic Publications:** Regularly publish findings in high-impact journals and conferences focused on AI, blockchain, and data security.
- **Open-Source Contributions:** Release software tools and frameworks developed during the research as open-source to facilitate community engagement and further research.
- **Workshops and Conferences:** Organize and participate in workshops and conferences to present research progress and gather feedback from peers.
- **Industry Collaborations:** Partner with industry stakeholders to pilot real-world applications of the research and promote technology transfer.

V. Supporting Infrastructure

- **Computational Resources:** Access to high-performance computing facilities and blockchain platforms provided by the Université de Paris 1 Panthéon-Sorbonne.
- **Software Tools:** Utilization of advanced software tools for AI development, blockchain implementation, and data analytics.
- **Collaboration Platforms:** Online collaboration tools for seamless communication and document sharing among the research team and external advisors.

VI. Timeline

- **Phase 1: Initial Research and Planning**, 6 months, Literature review, state-of-the-art analysis, and research proposal refinement.
- **Phase 2: Design and Modeling**, 12 months, Development of blockchain-based models and frameworks.
- **Phase 3: Implementation**, 12 months, Prototype development and preliminary testing.
- **Phase 4: Experimentation and Evaluation**, 9 months, Extensive testing, validation, and refinement of systems.
- **Phase 5: Dissemination and Finalization**, 9 months, Publication of results, software release, and thesis writing.

VII. Advisory Group Membership

- **Thesis Director:** Pr. Camille Salinesi, Professor at the Université de Paris 1 Panthéon-Sorbonne
- **Thesis Co-Director:** Dr. Nicolas Herbaut, Associate Professor at the Université de Paris 1 Panthéon-Sorbonne.

VIII. References

1) Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., G. Zelaya, C., Van Moorsel, A. (2020). "The relationship between trust in AI and trustworthy machine learning technologies." In Proceedings of the 2020 conference on fairness, accountability, and transparency, 272-283.
2) Mammen, P. M. (2021). "Federated learning: Opportunities and challenges." arXiv preprint arXiv:2101.05428.
3) Tolpegin, V., Truex, S., Gursoy, M. E., Liu, L. (2020). "Data poisoning attacks against federated learning systems." In Computer security–ESORICS 2020, 480-501.
4) Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., Zou, J. (2022). "Advances, challenges and opportunities in creating data for trustworthy AI." Nature Machine Intelligence, 4(8), 669-677.
5) Nassar, M., Salah, K., Rehman, M. H. ur, Svetinovic, D. (2020). "Blockchain for explainable and trustworthy artificial intelligence." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(1), e1340.
6) Six, N., Perrichon-Chrétien, A., Herbaut, N. (2022). "Saiaas: A blockchain-based solution for secure artificial intelligence as-a-service." In The international conference on deep learning, big data and blockchain (deep-BDB 2021), 67-74.
7) Xu, H., Qi, S., Qi, Y., Wei, W., Xiong, N. (2024). "Secure and lightweight blockchain-based truthful data trading for real-time vehicular crowdsensing." ACM Transactions on Embedded Computing Systems, 23(1), 1-31.
8) Jovanovic, Zorka, et al. "Robust integration of blockchain and explainable federated learning for automated credit scoring." Computer Networks 243 (2024): 110303.
9) Sachan, Swati, and Xi Liu. "Blockchain-based auditing of legal decisions supported by explainable AI and generative AI tools." Engineering Applications of Artificial Intelligence 129 (2024): 107666.
10) Chen, Hsin-Yuan, et al. "Integrating explainable artificial intelligence and blockchain to smart agriculture: Research prospects for decision making and improved security." Smart Agricultural Technology 6 (2023): 100350.