Blockchain for Trustworthy AI in Healthcare: A Systematic Review Methodology

Samer El Hajj Hassan
Computer Science Department
International University of Applied Sciences,
Berlin, Germany
samer.elhajjhassan@iu-study.org

***Abstract*— The integration of blockchain technology with artificial intelligence (AI) in healthcare aims to address critical challenges related to trustworthiness, privacy, and data integrity. This methodology section outlines the approach for a systematic literature review (SLR) focusing on blockchain-based solutions for enhancing trustworthy AI in healthcare. The review seeks to identify, evaluate, and synthesize existing research to understand how blockchain can support AI's trustworthiness across various healthcare applications.**

Keywords*— *Blockchain, AI, Healthcare, Trustworthy, Data-Privacy

I. INTRODUCTION

The healthcare industry increasingly relies on AI to improve patient outcomes, optimize operational efficiency, and facilitate medical research [1]. However, the trustworthiness of AI systems remains a critical concern, particularly regarding data integrity, privacy, and transparency [2][3][4]. Blockchain technology, with its inherent properties of immutability, decentralization, and transparency, offers promising solutions to address these concerns [5][6][7]. This systematic literature review will explore the current state of research on the use of blockchain to support trustworthy AI in healthcare.

II. MOTIVATION

The motivation behind this systematic literature review (SLR) stems from the need to ensure that AI systems in healthcare are not only effective but also trustworthy. Trustworthy AI in healthcare is underpinned by three key pillars: Trustworthy Machine Learning [8][9], Trustworthy Inference (often linked with Explainable AI, or XAI) [10][11], and Trustworthy Data Management [12]. These elements are fundamental to the development of AI systems that are ethical, reliable, and capable of making decisions in a transparent and understandable manner, thus enhancing user trust.

Blockchain technology, with its properties, offers promising solutions to support these pillars of trustworthy AI. Specifically, blockchain enhances Trustworthy Data Management by ensuring high-quality, reliable and trustworthy data, which is essential for AI training. It can enhance data integrity, ensure privacy-preserving machine learning through federated learning [8], and provide a robust foundation for explainable AI. This review aims to explore how blockchain technology can support the development and deployment of trustworthy AI systems in healthcare, addressing critical issues such as data preparation, privacy, and transparency. By systematically reviewing the literature, we aim to provide insights into current applications, challenges, and future research directions in this interdisciplinary field.

III. METHODOLOGY

This methodology section outlines the systematic approach employed to conduct a SLR on the integration of blockchain technology with AI to support trustworthiness in healthcare. The review follows a structured protocol designed to ensure comprehensive coverage and unbiased selection of relevant studies [13].

A. Research Questions

This SLR is guided by the following research questions (RQs):

RQ1: What are the use cases of integrating blockchain with AI for trustworthiness in healthcare?

RQ2: Among the identified use cases, which blockchain-based applications have been developed in alignment with AI and trustworthiness?

RQ3: What are the challenges and limitations of implementing blockchain for trustworthy AI in healthcare?

RQ4: What are the open research issues and the areas for future research in the integration of blockchain and AI for trustworthiness in healthcare?

Subsequently, these research questions will aim to address the following research inquiries (RIs):

RI1: How can blockchain technology be utilized to build a reliable data preparation pipeline for AI?

RI2: What are the methods to employ blockchain in order to foster trust in machine learning processes?

RI3: How can blockchain be used to ensure the interpretability of AI predictions?

RI4: What are the strategies to globally assemble trustworthy AI components in order to develop a fully reliable end-to-end AI solution?

B. Research Objective

The objective of this systematic literature review is to identify, analyze, and synthesize existing research on the integration of blockchain technology with AI to enhance trustworthiness in healthcare [14]. The review aims to provide a clear understanding of current developments, challenges, and future research directions in this field, including the examination how blockchain can support data preparation, privacy-preserving machine learning, and explainable AI (XAI) to foster credibility and trust in AI outcomes within the healthcare sector.

C. Research Strategy

A systematic mapping process provides a wide overview of the SLR. This approach helps map blockchain, AI, and trustworthiness in healthcare, ensuring comprehensive coverage of the research landscape. The process involves several steps [13]:

- Definition of RQs: Identify key questions to guide the review.

- Search Strategy Development: Formulate search queries and identify relevant databases.
- Study Selection: Apply inclusion and exclusion criteria to filter relevant studies.
- Data Extraction: Extract appropriate information from selected studies.
- Data Synthesis: Analyze and synthesize data to answer the RQs.

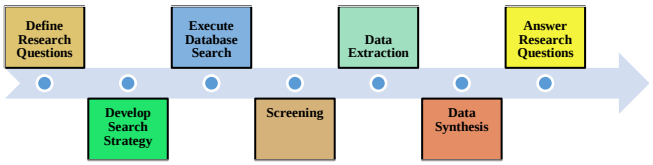

Fig. 1. Systematic Mapping Process

To identify relevant primary studies for our systematic literature review on blockchain, AI, and trustworthiness in healthcare, we will employ a comprehensive search strategy. This strategy includes the following steps [13]:

- Search Terms: We will use a combination of keywords to ensure a broad and inclusive search. Key terms include "blockchain," "AI," "Artificial Intelligence," "Machine Learning," "Healthcare," and "trustworthiness."
- Resources: Our search will cover a variety of resources, including major digital libraries such as Scopus academic database using a meta-search engine provided by MIAGE Scholar [13].
- Initial Mapping Study: An initial mapping study will be conducted to refine search terms and identify the most relevant resources. This preliminary review ensures to cover all relevant studies effectively.

D. Study Selection

The search query is formulated as follows:

```
(TITLE-ABS-KEY(blockchain) OR TITLE-ABS-KEY("Blockchain"))
AND (TITLE-ABS-KEY(AI) OR TITLE-ABS- KEY("Artificial
Intelligence") OR TITLE-ABS-KEY("Machine Learning") OR TITLE-
ABS-KEY("machine learning"))
AND (TITLE-ABS-KEY(Healthcare) OR TITLE-ABS-KEY(Health))
AND (TITLE-ABS-KEY(trustworthy) OR TITLE-ABS-KEY(trust) OR
TITLE-ABS-KEY(Trustworthy) OR TITLE-ABS-KEY(Trust))
AND PUBYEAR > 2020
AND DOCTYPE(ar)
```

This query ensures a comprehensive search for articles published after 2020 that focus on blockchain, AI, and trustworthiness in healthcare.

The initial screening involves:

- Accessibility: Filtering for Open Access articles.
- Excluding non-article document types (e.g., Overview, Conference papers, Book chapters, Magazine articles, Theses, Interview-based articles.
- Publication Year: Ensuring the publication year published after 2020.

E. Screening Selection

The screening selection process follows a multi-step approach to ensure the inclusion of relevant and high-quality studies:

1) Inclusion Criteria

The inclusion criteria for the SLR includes:

- Document Type: Peer-reviewed journal articles.
- Relevance: Articles must explicitly discuss the integration of blockchain with AI for trustworthiness aspects such as data integrity, privacy and explainability in healthcare.

2) Exclusion Criteria

Studies are excluded if they:

- Do not focus on the healthcare sector.
- Are not available as Open Access.
- Duplicated studies
- Eliminate studies based on quality evaluation criteria.

After applying these criteria, 43 out of 189 preliminary search Scopus research publications were selected (shown in **https://s.miage.dev/sbeaW**) for further review.

F. Data Extraction

Data extraction is a critical step where relevant information is systematically collected from each selected study [13]. This includes:

- Study Details: Title, authors, journal, publication year, DOI.
- Research Focus: Key research questions addressed by the study.
- Methodology: Study design, data sources, analytical methods.
- Findings: Key findings related to the use of blockchain for trustworthy AI in healthcare.
- Limitations: Any limitations identified by the authors.

G. Data Synthesis

Each study is assessed for quality using the following criteria:

- Article Quality Index: Prioritize studies from Q1 and Q2 journals.
- Number of Citations: Higher citation counts indicate greater impact.
- Journal Impact: Consider the impact factor and reputation of the publishing journal.

H. Findings

The reviewed literature identifies key use cases for blockchain in supporting trustworthy AI in healthcare, such as secure medical data management and decentralized clinical trials. Blockchain-based applications like federated learning frameworks and smart contracts for health insurance claims have been developed to enhance data security and transparency. However, challenges such as scalability, high computational costs, and regulatory compliance persist. Future research should address these issues and explore the integration of blockchain with emerging technologies to enhance AI trustworthiness in healthcare.

The findings will be documented comprehensively, highlighting current research, gaps, and future directions. This SLR aims to provide valuable insights, guiding future research and development in blockchain-supported trustworthy AI in healthcare, focusing on ethical, reliable, and efficient AI systems.

IV. REVIEW PROTOCOL

This protocol outlines the SLR process for integrating blockchain with AI to enhance trustworthiness in healthcare. It details the planning, execution, and dissemination phases, ensuring comprehensive and unbiased selection and synthesis of relevant studies. The protocol schema is outlined in Table 1.

TABLE 1. THE PROTOCOL FOR A SYSTEMATIC LITERATURE REVIEW

I. Planning *Scope of the Review*	**1. Background:** Justify the SLR need by highlighting literature gaps: Integrating blockchain with AI enhances trustworthiness, privacy, and data integrity in healthcare systems. **2. Research Questions:** Define guiding questions and goals (RQs and IQs). **3. Research Strategy:** • Search terms and keywords: (e.g. "blockchain," "AI," "Artificial Intelligence," "Machine Learning," "Healthcare," and "trustworthiness.") • Resources to be searched (digital libraries e.g. Scopus publication Database) • Initial Mapping Study: Conduct a preliminary review to refine search terms and identify relevant resources. **4. Study Selection Criteria:** • Inclusion Criteria: Peer-reviewed journal articles that discuss the integration of blockchain with AI for trustworthiness in healthcare, published after 2020, available as Open Access, English. • Exclusion Criteria: Non-healthcare focused studies, Duplicated. • Pilot Selection: Test the selection criteria on a subset of studies to ensure effectiveness. **5. Study Selection Procedures:** • Application: Apply inclusion and exclusion criteria during the initial screening of titles and abstracts. • Evaluation: Review full texts of potentially relevant studies. • Quality Check: Reevaluate uncertain studies to ensure they meet the criteria. • Specify the number of assessors. • Resolution: Outline the process for resolving disagreements (e.g., consensus or third assessor). • Final Selection: Include studies meeting all criteria.
II. Execution *Article Identification for further analysis*	**6. Study Quality Assessment**: • Checklists: Develop quality checklists based on article quality index, number of citations, and impact factor. • Assess each study's quality to ensure high standards. **7. Data Extraction Strategy from articles:** • Information Extraction: Collect relevant data including title, authors, year, findings, and relevance from each study. • Validation: Cross-check extracted data for accuracy and document any assumptions or inferences made.
III. Assimilation *Study Scope*	**8. Synthesis of the Extracted Data:** • Synthesis Strategy: Summarize findings from studies. • Meta-Analysis: Conduct, if applicable, a statistical combination technique of results. **9. Dissemination Strategy** • Publication: Submit to a peer-reviewed journal. • Conferences: Present findings at conferences. • Open Access: Ensure open access availability. **10. Project Timetable (Milestones and Deadlines)** • Weeks 1: Develop questions and search strategy. • Weeks 2: Screen titles and abstracts. • Weeks 3: Full-text review and selection. • Weeks 4: Data extraction and validation. • Weeks 5: Data synthesis and meta-analysis. • Weeks 6-7: Write and finalize manuscript. • Week 8: Submit for publication.

V. CONCLUSION

This SLR aims to provide a comprehensive understanding of how blockchain technology can support trustworthy AI in healthcare. By synthesizing existing research, identifying current applications, challenges, and future research directions, this review will contribute to developing secure, transparent, and reliable AI systems in the healthcare sector. The findings will inform stakeholders, including researchers, practitioners, and policymakers, about the potential and limitations of integrating blockchain with AI in healthcare.

REFERENCES

[1] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, *6*(2), 94.

[2] Yadav, N., Pandey, S., Gupta, A., Dudani, P., Gupta, S., & Rangarajan, K. (2023). Data Privacy in Healthcare: In the Era of Artificial Intelligence. Indian dermatology online journal, 14(6), 788–792. https://doi.org/10.4103/idoj.idoj_543_23.

[3] Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics*, *22*, 1-5.

[4] Williamson, S. M., & Prybutok, V. (2024). Balancing privacy and progress: A review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. *Applied Sciences*, *14*(2), 675. https://doi.org/10.3390/app14020675

[5] Kumar, P., Kumar, R., Aloqaily, M., & Islam, A. N. (2023). Explainable AI and blockchain for metaverse: A security, and privacy perspective. *IEEE Consumer Electronics Magazine*.

[6] Jovanovic, Z., Hou, Z., Biswas, K., & Muthukkumarasamy, V. (2024). Robust integration of blockchain and explainable federated learning for automated credit scoring. *Computer Networks*, *243*, 110303.

[7] Sachan, S., & Liu, X. (2024). Blockchain-based auditing of legal decisions supported by explainable AI and generative AI tools. *Engineering Applications of Artificial Intelligence*, *129*, 107666.

[8] Mammen, P. M. (2021). Federated learning: Opportunities and challenges. *arXiv preprint arXiv:2101.05428*.

[9] Tolpegin, V., Truex, S., Gursoy, M. E., & Liu, L. (2020). Data poisoning attacks against federated learning systems. In Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part I 25 (pp. 480-501). Springer International Publishing.

[10] Nassar, M., Salah, K., ur Rehman, M. H., & Svetinovic, D. (2020). Blockchain for explainable and trustworthy artificial intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(1), e1340.

[11] Rawal, A., McCoy, J., Rawat, D. B., Sadler, B. M., & Amant, R. S. (2021). Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence*, *3*(6), 852-866.

[12] 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.

[13] Kitchenham, B. A. & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering* (EBSE 2007-001). Keele University and Durham University Joint Report .

[14] Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & Van Moorsel, A. (2020, January). The relationship between trust in AI and trustworthy machine learning technologies. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 272-283).