A Systematic Literature Review on Enterprise Blockchain Supporting Trustworthy AI

Yuntian DING
Université Paris 1 Panthéon-Sorbonne Bordeaux, France ding.yuntian98pro@gmail.com

***Abstract*—Artificial intelligence adoption is increasingly considered by industries attempting to enhance efficiency in their domains. However, doubts and even resistance towards its large-scale adoption persist. Improving the trustworthiness of AI and regulating and monitoring AI models remain significant challenges. Meanwhile, blockchain, known for its trust-providing capabilities, offers potential solutions worthy of exploration. A systematic literature review (SLR) is essential to explore the intersection of trustworthy AI and the blockchain technology. This paper describes the methodology for generating a systematic literature review (SLR) aiming to analyze existing research, identify gaps, and guide furture research orientation.**

***Index Terms*—Trust; Blockchain; AI**

I. Introduction

As AI technology becomes increasingly widely used, its ethicality and trustworthiness are also growing in importance. Enhancing AI trustworthiness involves improvements in three facets: the trust in machine learning (ML), the trust in interface (Explainable AI or XAI), and the trust in data management. These aspects have their significant challenges affecting their trustworthiness, hence the trustworthiness of AI, including issues like hallucination (this means the output contains false or misleading information), lack of transparency and the need for data security and privacy. Blockchain, a widely recognized trust maker, has features including decentralization, immutability, security and transparency. It is commonly belived that its development and integration offer an opportunity to address the challenges and foster trustworthy AI. Therefore, this paper aims to conduct a systematic literature review on the potential solutions offered by blockchain to foster trustworthy AI.

II. Motivation

A comprehensive systematic literature review that can cross-analyze the two AI and Blockchain, and compare blockchain's impacts with more mature regulatory technologies can provide valuable insights into increasing the trustworthiness of AI. By analyzing and classifying existing research efforts, the SLR aims to summarize previous work, identifying current contributions and research gaps, thereby positioning our research proposal and evaluating our potential contributions.

III. Review Process

This systematic literature review (SLR) is conducted based on Kitchenham's Guidelines for performing Systematic Literature Reviews in Software Engineering [1]. According to these guidelines, the review process is divided into three steps :

- Review Planning: Define the needs of this SLR to generate an unbiased and complete overview of existing information in the area of interest. It is crucial to rigorously define the research questions (RQs) during this process, as these will directly influence on the completeness and relevance of this review.
- Review Conducting: Develop a search strategy and selection criteria then summarize and analyse the related studies to address the research questions.
- Review Reporting: Obtain the final output of our SLR, evaluate it and discuss the dissemination strategy to deliver this output to the interested public. This paper does not discuss this stage in detail, as it focuses on the methodology applied to undertake an SLR.

IV. Review Protocol

In order to undertake the SLR, we defined a protocol including the following components:

- Research Questions
- Sources and search strategy
- Inclusion and Exclusion Criteria
- Quality assessment (achieved after the first selection based on the inclusion and exclusion criteria, allowing further refinement of the selection)
- Data extraction strategy (defining how to collect information from the selected studies)
- Data synthesis (summarising the information of the final dataset defined after the data extraction step)

To ensure our SLR won't add further bias, it is improtant to check the elements listed in this protocol periodically.

V. Review Planning

The essential and initial step in this stage is to define research questions. Rigorously defined research questions are crucial for targeting the research interest and ensuring the quality of the systematic literature review.

A. Research Questions

To ensure the effectiveness of the research questions, the PICO (Population, Intervention, Comparison, Outcome) criteria is applied to frame them :

- Population : Trustworthy AI and its three facets (trustworthy machine learning, trustworthy interface (Explainable AI or XAI), trustworthy data management)
- Intervention : Integration with blockchain technology

- Comparison : AI optimized or controled by other existing methods on the aspect of trust
- Outcome : Evaluate the improvement in the trustworthiness of AI by integrating with the blockchain technology

The research questions are as the follow :

1) ***RQ1****:* Which Blockchain's capabilities can enhance the trustworthiness of AI, percisely on the trustworthiness of ML, the trustworthiness of interface and the trustworthiness of data management, compared to other existing regulating and securing systems ?

2) ***RQ2****:* What evidence and metrics are needed to fill the gap between empirical insights and theoretical results at the intersection of blockchain application and trustworthy AI ?

3) ***RQ3****:* What are the concerns or remaining challenges in using blockchain technology to address the trustworthiness of AI in the current research background ?

VI. Review Conducting

In this section, the methodology for searching and collecting information will clearly defined. Therefore, the following subsections will include source and search strategy, inclusion and exclusion criteria, quality assessment and data extraction strategy in accordance with our previouly defined review protocol.

A. Sources and Search Strategy

The following electronic sources will be used as our database:

- IEEExplore
- ACM Digital library
- Scopus
- Google Scholar
- SpringLink (for journals)

To formulate the search string, keywords are first extracted from our research questions and classifed into three categories : Population, Intervention and Outcome (Table I).

TABLE I
Key Words Identification and Classification

RQ1	
Population	AI OR Artificial Intelligence
Intervention	Blockchain
Outcome	trustworthy OR ethical OR trust
RQ2	
Population	AI OR Artificial Intelligence
Intervention	Blockchain
Outcome	Metrics OR Evidence OR Proof
RQ3	
Population	AI OR Artificial Intelligence
Intervention	Blockchain
Outcome	challenges OR pitfalls

The search strings will be generated by combining these keywords using Boolean operators (AND,OR). Note that these search strings will vary in accordance of any changes of the review protocol.

1) Examples of Search Strings:

- RQ1:
  - (TITLE-ABS-KEY("AI") OR TITLE-ABS-KEY("Artificial Intelligence")) AND TITLE-ABS-KEY("Blockchain") AND (TITLE-ABS-KEY("trustworthy") OR TITLE-ABS-KEY("ethical") OR TITLE-ABS-KEY("trust"))
- RQ2:
  - (TITLE-ABS-KEY("Trustworthy AI") OR TITLE-ABS-KEY("Trustworthy Artificial Intelligence")) AND TITLE-ABS-KEY("Blockchain") AND (TITLE-ABS-KEY("metrics") OR TITLE-ABS-KEY("evidence") OR TITLE-ABS-KEY("proof"))
- RQ3:
  - (TITLE-ABS-KEY("AI") OR TITLE-ABS-KEY("Artificial Intelligence")) AND TITLE-ABS-KEY("Blockchain") AND (TITLE-ABS-KEY("challenges") OR TITLE-ABS-KEY("pitfalls"))

After collecting the resources, snowballing will be done to apply the inclusion and exclusion criteria, followed by the quality assessment. Then, the studies in the final dataset will be read attentively, and a data extraction strategy will be applied to collect useful related information.

B. Inclusion and Exclusion Criteria

This is the selection criteria applied to primary studies since the studies resulting from our search string may be completely irrelevant to our research questions.

1) Inclusion Criteria:

- Papers focus on blockchain-related solutions to improve the trustworthiness of AI
- Papers mention the potential use of blockchain to improve the trustworthiness of AI even if they focus on other improvement solutions.
- Papers discuss blockchain adoption as trust-enhancing solution in other applications, they can provide metrics and evidence.

2) Exclusion Criteria:

- Secondary or tertiary studies.
- Duplicated studies.
- Non-peer-reviewed studies.
- Papers not written in English.
- Papers do not explore the aspect of trust.
- Papers unrelated to AI nor blockchain.
- Papers do not provide any metrics or evidence to support their conclusions.

C. Quality Assessment

To enhance the quality of our SLR. The further selection will be based on a quality checklist, ensuring the inclusion of high-related and high-quality studies that address fostering trustworthy AI through blockchain technology. The checklist should be developed by considering the following factors [1]:

- Bias: Selection bias, performance bias, measurement bias and attrition bias
- Internal Validity: Results fairness and validity inside the study.
- External Validty: Results maintains its fairness and validity even in a geralised study case or applicated outside the study.

D. Data Extraction Strategy

A well-defined data extraction strategy is necessary to efficiently identify and classify related information. A data collection form will be generated for each selected study. It should contain the following information [1] :

- Data Extractor: Person who execute the process
- Data Checker: Person who check the output
- Dates: Execution date and/or check date
- Title, authors, publication details
- Research Interest
  - Application Domain
  - Contextual information
  - Motivation
- Study Design and Methodologies
  - Method Selection Process
  - Research Questions
  - Models: Theoretical frameworks
  - Inputs: Metrics used in the study for process execution
- Study Results
  - Variance: Variability or dispersion listed
  - Accuracy: Degree of fairness and precision of the results
  - Contributions
  - Limitations: Potential contraints listed in the study
  - Future work or Challenges
- Quality Assessment
  - Research Bias: Any biases that may have impacts on the reliability and the validity of findings
  - Revelance to our research questions: Evalution of the study's alignmeent with our research questions

To ensure the effectiveness, a cross-check with other people would be useful, and a PhD student can consider test-retest process to check the output of data extraction process [1].

E. Data Synthesis

Once all the selected paper have been read, and the data extraction has been reviewed, we can move to the stage of summarization to position our work and find our potential contributions based on the results. This step may include the following activities [1]:

- Descriptive Synthesis: Use the data colloction form, identify the similarities and the differences between studies regarding how blockchain enhences AI trustworthiness.
- Quantitative Synthesis: Define and apply quantitative statistical methods and applying on the selected studies (Meta-analysis), then analyze the aggregated insights and choose a presentation mechanism to illustrate the results.

VII. Review reporting

This step is achieved after having the final results of the systematic literature review. In this stage, we need to consider the following tasks [1]:

- Specifying the Dissemination Strategy
- Formatting the Main Systematic Review Report
- Evaluating Systematic Review Reports: a peer review is expected at this stage.

VIII. Conclusion

This paper lists and presents essential methodologies and protocol to conduct an SLR for the three stages of review process. Note that this is the draft of the SLR, and the process will evolve along with a deeper understanding of the topic over time. We intent to better understand the current challenges and opportunies at the intersection between blockchain technology and trustworthy AI, this systematic literature review should guide our future research orientation.

Reference

[1] S. Keele and others, "Guidelines for performing systematic literature reviews in software engineering." Technical report, ver. 2.3 ebse technical report. ebse, 2007.