NLP-BERT AND OPTIMIZATION OF EFFICIENCY-SECURITY IN BLOCKCHAIN-ACCOUNTING SMART CONTRACTS

Totok Dewayanto[1)]

Abstract: In the evolving digital era, blockchain-based accounting systems leveraging smart contracts are increasingly crucial. Blockchain, known for its transparency and security, faces challenges in the efficiency and security of smart contracts, necessitating innovative solutions. This research explores the potential of Natural Language Processing (NLP) technology such as BERT to enhance these aspects within smart contracts. Grounded in blockchain and smart contract theories, the study investigates BERT's use in detecting financial fraud and improving transaction security. The research methodology employs Bibliometric Analysis with a PICO approach to formulate research questions and gather relevant data. Bibliometric analysis provides deeper insights into current developments and challenges in this field. The analysis reveals trends in document production, author collaborations, and keywords related to blockchain and BERT. Co-occurrence networks and thematic maps highlight the significance and relevance of these terms in the context of blockchain and smart contracts. In conclusion, this study makes a significant contribution to understanding how NLP technology can be optimized to enhance the efficiency and security of smart contracts in blockchain-based accounting systems. Further research is essential to develop BERT optimization in this context.

Keywords: Blockchain-Accounting, Smart Contracts, NLP-BERT, Efficiency and Security

AFFILIATION:
Diponegory University[1)]

CORRESPONDENCE:
totokdewayanto@lecturer.undip.ac.id[1)]

ARTICLE HISTORY
Received:
March 13, 2024

Revised:
April 2, 2024

Accepted:
May 26, 2024

THIS ARTICLE IS AVAILABLE IN:
http://ejournal.stiepena.ac.id/index.php/fe

INTRODUCTION

In today's digital era, blockchain-based accounting systems utilizing smart contracts are increasingly pivotal. Blockchain, renowned for its transparency and security, has revolutionized how transactions are conducted and data is stored. However, despite the manifold advantages offered by blockchain, several challenges persist. One such challenge is the efficiency and security of smart contracts. Smart contracts are computer programs executing instructions based on specific conditions, integral to numerous blockchain applications. Yet, they can also be susceptible to attacks and misuse if not designed and implemented correctly (Dai & Vasarhelyi, 2017).

Natural Language Processing (NLP) technologies like BERT have shown significant potential in enhancing the efficiency and security of smart contracts. BERT, Bidirectional Encoder Representations from Transformers, is a language model that has transformed how we understand and use text across various applications, including smart contracts. However, the implementation and optimization of BERT in this context remain relatively new and underexplored areas of research (Devlin et al., 2018).

Several studies have explored the optimization and implementation of BERT in smart contracts. For instance, research by Zhang et al. (2023) delved into the application of NLP technology for precise financial fraud detection, focusing on the implementation and optimization of the FinChain-BERT model. This study demonstrated that with proper optimization and implementation, BERT can yield excellent results in financial fraud detection.

Research Gap: Despite several studies on the application of BERT in accounting and blockchain, there appears to be room for further exploration on how BERT can be optimized to enhance the efficiency and security of smart contracts in blockchain-based accounting systems. This indicates a significant research gap in this area, necessitating further research to fill this void.

Research Question: Based on this research gap, the primary research question that arises is: "What is the influence of optimizing BERT implementation on the efficiency and security of smart contracts in blockchain-based accounting systems?" This question aims to evaluate the direct impact of BERT optimization and implementation on smart contracts in the context of blockchain accounting.

Research Objective: The objective of this study is to answer this research question by conducting a bibliometric analysis of previous research on the optimization of BERT implementation in smart contracts for blockchain accounting. By doing so, this research hopes to provide new and valuable insights into how BERT can be optimally implemented in smart contracts to enhance their efficiency and security.

THEORETICAL BACKGROUND

In this context, the primary theories used are blockchain and smart contracts. Blockchain is a technology that enables secure and transparent transactions without the need for a central authority (Nakamoto, 2008). Smart contracts are computer programs that execute instructions based on specific conditions and are integral to many blockchain applications (Buterin, 2013).

Previous research has shown the significant potential of NLP technologies like BERT in improving the efficiency and security of smart contracts (Devlin et al., 2018). Research by Zhang et al. (2023) also indicates that with proper optimization and implementation, BERT can deliver excellent results in financial fraud detection.

The conceptual framework of this research is based on the idea that optimizing BERT implementation can enhance the efficiency and security of smart contracts in blockchain-based accounting systems. This involves a bibliometric analysis of previous research on the optimization of BERT implementation in smart contracts for blockchain accounting.

In this study, we will use this framework to address the research question: "What is the influence of optimizing BERT implementation on the efficiency and security of smart contracts in blockchain-based accounting systems?". By answering this question, we aim to provide new and valuable insights into how BERT can be optimally implemented in smart contracts to enhance their efficiency and security.

METHOD

Based on the method provided, here's a breakdown of how the study would proceed:

PICO Framework:

- Population (P): Implementation of BERT in smart contracts for blockchain accounting.
- Intervention (I): Optimization of BERT implementation in smart contracts.
- Comparison (C): Studies on the implementation of BERT in smart contracts for blockchain accounting without bibliometric analysis.

- Outcome (O): Better understanding of developments and changes in this research field over time, challenges faced, and the impact of optimizing BERT implementation on the efficiency and security of smart contracts.

Search String:

TITLE-ABS-KEY (("blockchain accounting" OR "smart contracts") AND (ilm OR nlp OR bert) AND (efficient OR security))

Study Selection:

PRISMA

1. Include 32
2. Exclude (-)

32 document result

ANALYSIS

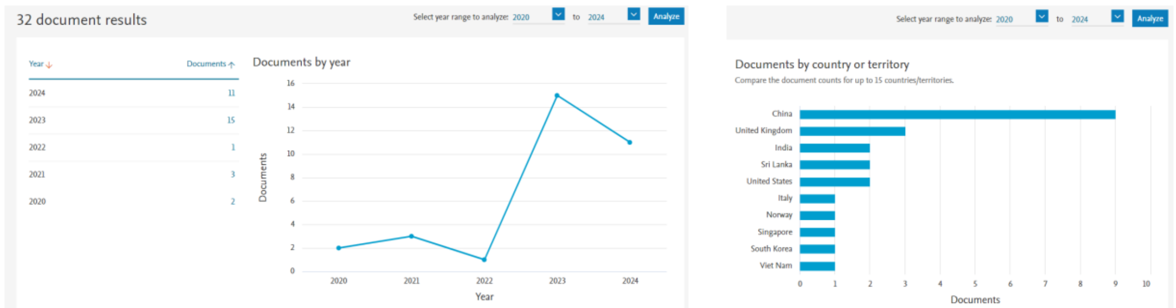

Figure 1. Document Results

Here is the translation of the further explanation about the Figure1. Country or Region: The list of countries or regions displayed includes China, England, India, Sri Lanka, the United States, Italy, Norway, Singapore, South Korea, and Vietnam.

Number of Documents: The bars on the diagram represent the number of documents for each country or region within the selected time frame. China has the highest number of documents, with approximately 10 documents, followed by England and India with around 6 and 5 documents respectively. Other countries have progressively fewer documents, as indicated by the shorter bars on the graph.

This image is interesting because it provides a visual representation of data that can be used to compare the number of documents across various countries or regions over a specific period of time. This could be highly useful for analyzing trends in document production or distribution among these regions within the given years.

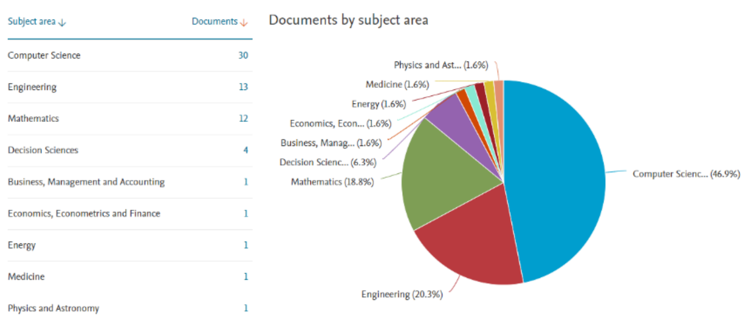

Figure 2. Document by Subject Area

The image above is a pie chart depicting the distribution of documents based on subject areas. The largest segment is Computer Science (46.9%), followed by Engineering (23.9%), Medicine and Arts (4.6% each), Economics, Energy, Business, Management, and Accounting (3.6% each), Decision Science (3.3%), Mathematics (3.3%), Social Sciences and Finance (2.8% each), and

Physics and Astronomy (1.9% each). This illustrates which areas have more material available, potentially indicating higher research activity or interest in these fields.

Table 1. Main Information

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2020:2024
Sources (Journals, Books, etc)	24
Documents	32
Annual Growth Rate %	53,14
Document Average Age	1,06
Average citations per doc	1,875
References	576
DOCUMENT CONTENTS	
Keywords Plus (ID)	177
Author's Keywords (DE)	143
AUTHORS	
Authors	77
Authors of single-authored docs	2
AUTHORS COLLABORATION	
Single-authored docs	2
Co-Authors per Doc	2,47
International co-authorships %	9,375
DOCUMENT TYPES	
article	7
book	1
book chapter	1
conference paper	11
conference review	12

Here is the translation of the main information about the data:

- Timespan: The data covers the period from 2020 to 2024.
- Sources (Journals, Books, etc): There are 24 sources of data used, including journals, books, and others.
- Documents: There are 32 documents in the dataset.
- Annual Growth Rate %: The annual growth rate of documents is 53.14%.
- Document Average Age: The average age of documents is 1.06 years.
- Average Citations per Document: On average, each document is cited 1.875 times.
- References: There are 576 references in the data.

Document Contents:
- Keywords Plus (ID): There are 177 additional keywords in the data.
- Author’s Keywords (DE): There are 143 keywords provided by the authors in the data.

Authors:
- Authors: There are 77 authors in the data.
- Authors of Single-authored Docs: There are 2 authors who wrote documents individually.

Collaboration among Authors:
- Single-authored Docs: There are 2 documents written by a single author.
- Co-Authors per Document:** On average, there are 2.47 authors per document.
- International Co-authorships %:** 9.375% of the documents have authors from more than one country.

Document Types:
- Article: There are 7 articles.
- Book: There is 1 book.

- Book Chapter: There is 1 book chapter.
- Conference Paper: There are 11 conference papers.
- Conference Review: There are 12 conference reviews.

Table 2. Co-occurrence Network

Node	Cluster	Betweenness	Closeness	PageRank
block-chain	1	128,8886012	0,035714286	0,140760948
blockchain	1	47,26866771	0,03030303	0,10339819
smart contract	1	57,67216785	0,03125	0,115227676
deep learning	1	5,459678109	0,024390244	0,065521488
vulnerability detection	1	14,43124241	0,026315789	0,072263332
ethereum	1	26,76049956	0,022222222	0,046580859
semantics	1	2,499546106	0,022222222	0,044238686
codes (symbols)	1	0,153846154	0,020833333	0,031634284
language model	1	0	0,02	0,022504383
losses	1	0,503113553	0,02173913	0,028426128
network security	1	0	0,020408163	0,021295385
static analysis	1	0,076923077	0,020833333	0,025135242
cloning	1	0	0,018867925	0,010268143
decentralised	1	0	0,019607843	0,015656116
detection methods	1	0	0,020408163	0,017848869
ecosystems	1	0	0,018518519	0,007849785
financial loss	1	0	0,020408163	0,018124781
network coding	1	0	0,020408163	0,020049064
security vulnerabilities	1	0	0,019230769	0,012709489
source codes	1	0	0,02	0,017488193
natural language processing systems	2	1,285714286	0,020833333	0,031199578
language processing	2	0	0,02	0,020057209
natural language processing	2	0	0,02	0,020057209
natural languages	2	0	0,02	0,020057209
pre-training	3	0	0,020408163	0,02118667
pre-training technique	3	0	0,020408163	0,02118667
training techniques	3	0	0,020408163	0,02118667
classification (of information)	4	0	0,014084507	0,008087745

Here is the translation of the explanation for each column in the co-occurrence network table:

1. **Node**: This is the entity or 'node' in the network. In this context, a node represents a keyword or concept.
2. **Cluster**: This indicates the group or 'cluster' where the node belongs. Nodes within the same cluster are typically more closely related to each other compared to nodes in different clusters.
3. **Betweenness**: This is a measure of centrality in the network based on shortest paths. Nodes with high betweenness values often act as intermediaries or bridges in the network.
4. **Closeness**: This is a measure of how 'central' a node is within the network. Nodes with high closeness values are usually closer to all other nodes in the network.
5. **PageRank**: This is a measure of the importance of a node in the network. PageRank is an algorithm used by Google Search to rank web pages in their search results.

For example, "block-chain" is a node in Cluster 1 with a betweenness value of 128.8886012, closeness value of 0.035714286, and PageRank value of 0.140760948. This indicates that "block-chain" is an important node in this network.

Table 3. Thematic Map

Occu-rrences	Words	Cluster	Cluster_ Label	btw_ centrality	clos_ centrality	pagerank_ centrality
15	block-chain	1	block-chain	4446,338791	0,004672897	0,048576634
12	blockchain	1	block-chain	2756,973564	0,004032258	0,038675716
12	smart contract	1	block-chain	2063,366003	0,003759398	0,037313228
8	deep learning	1	block-chain	949,0074049	0,003496503	0,025899567
7	vulnerability detection	1	block-chain	320,9516055	0,002849003	0,021431145
6	ethereum	1	block-chain	651,7099179	0,002849003	0,019077973
5	semantics	1	block-chain	361,885163	0,003289474	0,016264923
4	codes (symbols)	1	block-chain	185,3932704	0,002849003	0,012916664
4	language model	1	block-chain	576,3415948	0,00330033	0,013777364
3	losses	1	block-chain	78,56981806	0,002427184	0,010348377

Based on the provided information, here's the clear and structured translation:

Thematic Map Table Explanation

The Thematic Map table illustrates the distribution of keywords or concepts related to blockchain, smart contracts, and BERT. It shows the frequency of occurrence of these keywords or concepts within the selected documents.

In this table, several columns provide key information about each keyword or concept. Some of the most important columns include:

- Occurrences: The number of times each term or concept appears in the dataset.
- Words: The total count of words related to each term or concept.
- Cluster: The group or cluster to which each term or concept belongs. In this case, all terms are within the "block-chain" cluster.
- Cluster_Label: The label or name assigned to each cluster.
- btw_centrality: Betweenness centrality metric for each term or concept. This metric measures how often each term or concept appears on the shortest path between other terms or concepts.
- clos_centrality: Closeness centrality metric for each term or concept. This metric measures how close each term or concept is to all other terms or concepts in the network.
- pagerank_centrality: PageRank centrality metric for each term or concept. This metric measures the importance or relevance of each term or concept in the network.

The numbers in the table represent the values for each metric for each term or concept. For example, the term "block-chain" has 15 occurrences, 4446 related words, betweenness centrality of 0.004672897, etc.

The table indicates that terms such as "block-chain," "blockchain," and "smart contract" are highly central in the network, with high betweenness and closeness centrality values. This suggests these terms are important and relevant in the context of blockchain and smart contracts. Terms like "deep learning" and "language model" also have relatively high betweenness and closeness centrality values, indicating they are important concepts in the network.

On the other hand, terms like "vulnerability detection" and "ethereum" have lower betweenness and closeness centrality values, indicating they are less central in the network. Overall, the table provides insights into the importance and relevance of each term or concept in the context of blockchain, smart contracts, and BERT.

CONCLUSION

The conclusion drawn from the article provided is that blockchain-based accounting systems utilizing smart contracts are increasingly crucial in today's digital era. While blockchain offers numerous advantages, there are challenges that need to be addressed, such as the efficiency and security of smart contracts. Natural Language Processing (NLP) technologies like BERT have shown significant potential in enhancing the efficiency and security of smart contracts. Several studies indicate that with proper optimization and implementation, BERT can deliver excellent results in detecting financial fraud. Although there have been previous studies on the application of BERT in accounting and blockchain, there is still room for further research on how BERT can be optimized to enhance the efficiency and security of smart contracts in blockchain-based accounting systems.

This research aims to answer the research question on how the optimization of BERT implementation influences the efficiency and security of smart contracts in blockchain-based accounting systems. By conducting bibliometric analysis of previous research, this study hopes to provide new insights into how BERT can be optimized to improve the efficiency and security of smart contracts.

The research methodology employed the PICO (Population, Intervention, Comparison, Outcome) approach to formulate the research question. Subsequently, a specific search string was used to retrieve relevant data. Based on the data obtained, bibliometric analysis was conducted to gain a better understanding of the developments and changes in this research field.

Furthermore, additional information about the analyzed data includes the timespan, number of documents, annual growth rate, average document age, average citations per document, and types of documents in the dataset. Additionally, there is an analysis of author collaboration, document types, and the types of approaches used in the research.

The results of the co-occurrence network analysis highlight various related keywords or concepts within the network, while the thematic map provides an overview of the distribution of identified keywords in the analysis.

This research contributes to understanding how NLP technologies such as BERT can be optimized in the implementation of smart contracts to enhance efficiency and security in blockchain-based accounting systems.

REFERENCES

Dai, J., & Vasarhelyi, M. A. (2017). Toward blockchain-based accounting and assurance. Journal of Information Systems, 31(3), 5-21.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Zhang, Y., Xu, J., Chen, H., & Zhu, X. (2023). Optimization Strategies for BERT-Based Named Entity Recognition. In Advances in Knowledge Discovery and Data Mining (pp. 96-108). Springer, Cham.

Buterin, V. (2013). Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform. Ethereum.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin.org.

Dai, J., & Vasarhelyi, M. A. (2017). Toward blockchain-based accounting and assurance. Journal of Information Systems, 31(3), 5-21.

Zhang, Y., Xu, J., Chen, H., & Zhu, X. (2023). Optimization Strategies for BERT-Based Named Entity Recognition. In Advances in Knowledge Discovery and Data Mining (pp. 96-108). Springer, Cham.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin.org.