electronics

MDPI

Article

A Blockchain-Based Detection and Control System for Model-Generated False Information

Chenlei Liu [1,2,3], Yuhua Xu [1,2,3], Bing Hu [1,2,3] and Zhixin Sun [1,2,3,*]

[1] Key Laboratory of Broadband Wireless Communication and Sensor Network Technology (Ministry of Education), Nanjing University of Posts and Telecommunications, New Mofan Road No. 66, Nanjing 210003, China; 2019070270@njupt.edu.cn (C.L.); xuyh@njupt.edu.cn (Y.X.); hubing@njupt.edu.cn (B.H.)
[2] Post Big Data Technology and Application Engineering Research Center of Jiangsu Province, Nanjing University of Posts and Telecommunications, New Mofan Road No. 66, Nanjing 210003, China
[3] Post Industry Technology Research and Development Center of the State Posts Bureau (Internet of Things Technology), Nanjing University of Posts and Telecommunications, New Mofan Road No. 66, Nanjing 210003, China
* Correspondence: sunzx@njupt.edu.cn

Abstract: In the digital age, spreading false information has a far-reaching impact on various areas, such as society, politics, and the economy. With the popularization of applications of text generation models, the cost of producing false information has significantly decreased, making it challenging for human beings to screen it. Therefore, research on detection screening and early warning control for model-generated false information becomes particularly important. In this paper, we propose a model-generated false information detection and control system based on blockchain. Firstly, we design a model-generated false information detection method combining model-generated text discrimination based on a self-attention network and text similarity detection based on a twin network. Secondly, we construct a blockchain-based model-generated false information control and traceability system. It utilizes the proposed detection algorithm to provide early warning and control of model-generated false information involving important and sensitive events before social network release. For information judged to be model-generated false, the stored data on the blockchain is utilized to track and trace the publisher. Ultimately, experimental tests prove that the proposed detection method improves the accuracy of false information detection. In addition, the operational efficiency of the prototype system can meet quality of service requirements.

Keywords: blockchain; model-generated false information; information detection; traceability

check for updates

Citation: Liu, C.; Xu, Y.; Hu, B.; Sun, Z. A Blockchain-Based Detection and Control System for Model-Generated False Information. *Electronics* **2024**, *13*, 2984. https://doi.org/10.3390/electronics13152984

Academic Editors: Cheonshik Kim and Aryya Gangopadhyay

Received: 15 June 2024
Revised: 25 July 2024
Accepted: 26 July 2024
Published: 29 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the rapid development of the Internet and diversified social platforms, network communication modes have broken through the limitations of traditional models. The spread of digital information has become faster and faster, and the scope of dissemination has become broader, deepening the degree of information sharing and increasing the amount and impact of daily online false information [1]. In addition, due to the wide popularity of generative modeling (e.g., GPT-4) applications, the cost of generating high-quality text has significantly reduced, making fabricating false information easier. These models can generate content virtually indistinguishable from descriptions of real people, thereby spreading a large amount of false information on the Web. This information may include fake news, fake personal comments or opinions, or even complete descriptions of fake people or events. Due to the high quality, natural tone, and rich details of this model-generated content, it is difficult for ordinary users to recognize its authenticity, thus increasing the risk of confusing and misleading information, which can cause a severe social trust crisis [2]. Although not all model-generated information is untrue, it may be

maliciously utilized in some cases to mislead the public, influence public opinion, or even interfere with the usual social order. Therefore, detecting and identifying model-generated false information and implementing early warning control is an important component of false information governance research.

Deep learning methods can learn and recognize patterns from large datasets by simulating how the human brain processes information, thus effectively identifying and filtering model-generated false information. Compared to traditional manual review methods, deep learning provides an efficient and automated solution for security detection that needs to be geared toward massive amounts of information and has become a significant method for network information security detection. For example, Google's Jigsaw project utilizes deep learning techniques to identify phishing and fraudulent information, significantly improving the accuracy and efficiency of detection [3]. In addition, social media platforms such as Twitter [4] and Facebook [5] also utilize deep learning models to identify and flag suspicious content and reduce the spread of false information. In addition, with the continuous development of artificial intelligence and its related technology fields, many research works have explored novel technical methods for false information prevention and control [6], including graph analysis [7], knowledge interaction [8], and interpretable learning [1].

However, there are still deficiencies in the existing technical research:

- Due to the low cost of producing and disseminating model-generated false information, it can easily generate many retweets. It is difficult to trace the responsibility to the individual or even the source, supervise the problem before it occurs, and trace the problem after it occurs.
- The same or highly similar information may exist in false information generated by models. Repeated detection increases costs and reduces efficiency, resulting in evidence collection relying on traditional methods and delayed clarification. There needs to be more clarity in balancing timely and comprehensive false information detection and evidence-locking.

Therefore, studying a network information monitoring system that can perform model-generated false information detection in advance and support tracking the source of this false information release is of great significance and value.

Blockchain, a new distributed data ledger technology [9], ensures on-chain data storage with strong decentralization, tamper-resistance, and traceability due to its special network-wide node participation in data consensus and multi-redundant backup mechanism. Currently, blockchain is regarded as an important data storage foundation and is widely used in application scenarios such as digital copyright protection [10], supply chain information traceability [11], data forensics for the Internet of Vehicles (IoV) [12], and as a multimedia repository for the Internet of Things (IoT) [13]. The advantages of blockchain in data storage security and information traceability make integrating network information control and blockchain innovation research feasible [14–16]. However, although the existing schemes can accurately trace the source of false information, they need further exploration in blocking false information before its release. They cannot block the potential dissemination hazards of false information before its release, which presents significant limitations in the governance of false information on the network, resulting in lower governance strength. Therefore, the application of blockchain in false information governance still needs to be studied in depth.

In this paper, we design an innovative model-generated false information detection and control system based on blockchain technology. The system relies on the inherent immutability of blockchain to ensure the reliability and transparency of the information traceability and forensic process. In order to further enhance the detection effectiveness of model-generated false information, this paper proposes a comprehensive detection method that skillfully integrates model-generated text recognition technology driven by a self-attention mechanism and text similarity analysis technology under the twin network architecture. Specifically, the model-generated text discrimination module aims to accurately

detect and identify model-generated text content. In contrast, the text similarity detection module deeply analyzes and compares the similarity between pieces of false information to achieve comprehensive identification and effective control of false information.

The main innovations of this paper are as follows:

1. We propose a blockchain-based model-generated false information detection and control system. The system utilizes blockchain technology to achieve information traceability and forensics by storing the important event information of the network on the chain in advance, ensuring it cannot be tampered with, to construct a key information base of important data.
2. We propose a false information detection method that combines model-generated text discrimination based on a self-attention network with text similarity detection based on a twin network to improve the detection accuracy of model-generated false information.
3. We analyze and validate the proposed method through experimental analysis, examining the performance of the system detection model using public datasets and testing the prototype system's quality of service performance.

This paper is structured as follows. Section 2 investigates the current state of research related to the content of this paper. Section 3 describes the functionality of the proposed system. Section 4 describes the detection algorithms and similarity detection algorithms used in this paper. Section 5 conducts simulation experiments to evaluate the detection performance and quality of service performance of the system proposed in this paper. Section 6 summarizes the entire paper and suggests directions for future work.

2. Related Works

False information detection is of great significance for maintaining the security of network public opinion and the stability of social order. Research on network false information detection has become an important direction in network security research, with machine learning methods widely applied to obtain the textual features of false information and predict its content. Shrivastava et al. [17] proposed a false information detection model based on a system of differential equations to detect and eliminate false information on online social platforms. Verma et al. [18] proposed a two-stage benchmark model called WELFake, which is based on the word embedding of linguistic features for fake news detection using machine learning classification. Guo et al. [19] proposed a fake news detection model for hybrid linguistic contexts using a multi-scale converter to fully capture the semantic information of the text and obtain a more resilient feature space for semantic features in hybrid languages by extracting a more productive feature hierarchy for the initial textual content. Balshetwar et al. [20] proposed a fake news detection solution that takes sentiment as an important feature and analyzes the correlation of multiple missing data variables and data features of information using a deep neural network classifier. Raza et al. [21] proposed a web-based false information detection model based on the transformer architecture that utilizes information from news articles and social context to detect false news.

With the growing demand for traceability and accountability at the source of false information release, blockchain and its key technologies are gradually being applied in researching false information protection due to their trustless, decentralized, and traceability characteristics. Dwivedi et al. [14] proposed an online social platform framework based on blockchain and digital watermarking to control the dissemination of false information. Chen et al. [15] proposed a blockchain-based false information prediction scheme for social platforms that uses a customized proof-of-authority consensus algorithm and a weighted ranking algorithm to determine the integrity of false information. Shahbazi et al. [16] proposed a false information detection system integrating blockchain and natural language processing to detect and predict social false information, using machine learning algorithms and a decentralized blockchain framework. It provides authoritative proof of digital content. Chen et al. [22] proposed an entropy-based blockchain incentive

mechanism to reduce the negative impact of the malicious behavior of false information reviewers on a group-based fake news prevention system, which determines the corresponding rewards and penalties through entropy as a measure of voting inconsistency. Sengupta et al. [23] proposed a private chain-based false information detection model that employs expert knowledge to detect information content and construct a dynamic weighted voting method oriented to expert credibility via blockchain to improve detection accuracy. Duvvuri et al. [24] combined deep learning with blockchain, enabling users to distinguish between legitimate and fraudulent information and ensuring that the results cannot be tampered with once published. Dhall et al. [25] proposed a web-based blockchain and key watermarking-based information publishing framework to curb the spread of false news by immediate backtracking and positive tracking through blockchain records once it is detected. In addition, malicious message propagation can be identified and curbed by observing transactions on the blockchain and checking the forwarding density and rate of specific original messages that exceed a threshold. Xiao et al. [26] proposed a fast fake news detection scheme based on a network computing framework, which utilizes the techniques of software-defined networking, edge computing, blockchain, and Bayesian networks, where a blockchain server accommodates traffic reports submitted by vehicles in the vehicular network, calculates the probability of the existence of a traffic event, determines the authenticity of the report's content, and provides a time-sensitive service for passing vehicles.

This paper focuses on machine learning-based model-generated false information detection and blockchain-based false information traceability capabilities, combining false detection and traceability control functions. Aiming at social news, policy documents, economic activities, and other important events that can attract public attention, we propose a blockchain-based model-generated false information detection and control system to realize pre-warning before the release of model-generated false information and traceability of the information after the release. This system aims to safeguard the authenticity and reliability of network information and reduce the hazards of model-generated false information in terms of its impact on public opinion and the frequency of its occurrence.

3. System Design

3.1. System Model

Aiming at the current problem of the proliferation of false information on various network platforms and the lack of unified information detection, public opinion control, and traceability methods, this paper proposes a blockchain-based false information detection and control system, whose system model is shown in Figure 1.

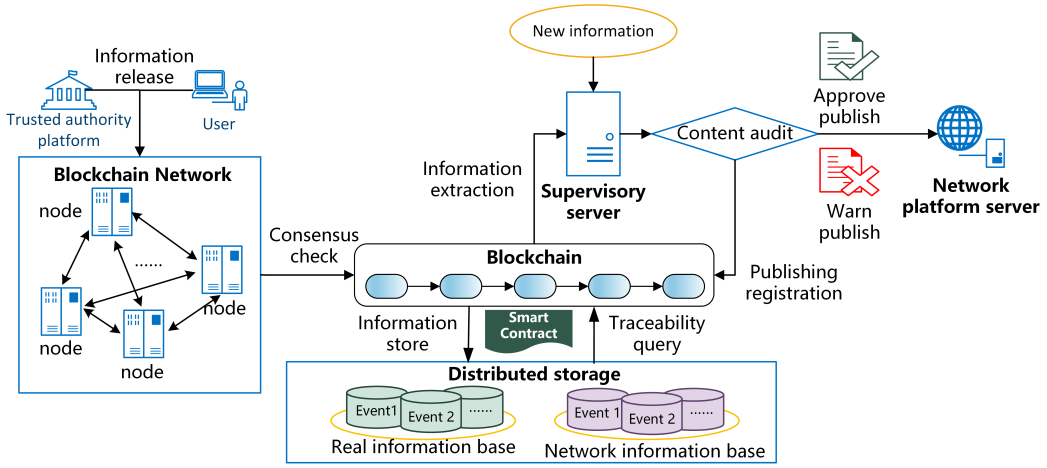


Figure 1. System model.

The core idea of the system lies in the following:

- We use blockchain to store information released by credible authorities (e.g., authoritative media, government agencies, etc.) and network users, registering them on the chain through the node consensus checking mechanism. The information released by credible authorities can be regarded as real information, which the detection model can use as benchmark data for false information detection.
- The supervisory server uses a false information detection algorithm and an information similarity comparison algorithm to detect and analyze information released by network users.
- Blockchain adopts on-chain and off-chain storage modes. It stores the information indexes and the information of the block on-chain, while the actual information content is stored off-chain through a distributed database classified according to the name of the information event to save blockchain storage space.

3.2. System Functions

The system integration proposed in this paper consists of two main functions—model-generated false information detection and traceability—to deal with today's increasingly sophisticated and insidious information deception threats. The information control module is dedicated to catching and intercepting potential model-generated false information threats before the release of information related to important events. It adopts text forgery detection technology to identify and block the dissemination of false information. The detection and traceability module focuses on after-the-fact traceability analysis, tracing the source of model-generated false information through blockchain-based traceability technology.

3.2.1. Model-Generated False Information Detection

The model-generated false information detection function includes information uploading, false information detection, text similarity calculation, and model-generated content warning. The function process is shown in Figure 2. The system first uploads and stores the data of the information release, including the event name, publisher, release time, content, etc. Subsequently, the system calls the detection interface of the functional layer, performs feature extraction on the collected data, and preprocesses the extracted features. Based on the extracted features, the system uses the trained binary self-attention model to classify the text and determine whether the text is model-generated content or standard human-generated content.

The system designed in this paper is for detecting false information content on social platforms for major social events. For the term "tweets" on social platforms, we believe that if model-generated content is identified, the information detection results can be directly considered false. Then, the detection results are registered on the chain and stored. If a human has written the content, it is necessary to detect similarity with the accurate information stored in the blockchain under the same event name. If the similarity exceeds the threshold, the information is released automatically. Otherwise, the detection results are uploaded to the blockchain as evidence, the release of the message is blocked, and the publisher is held accountable.

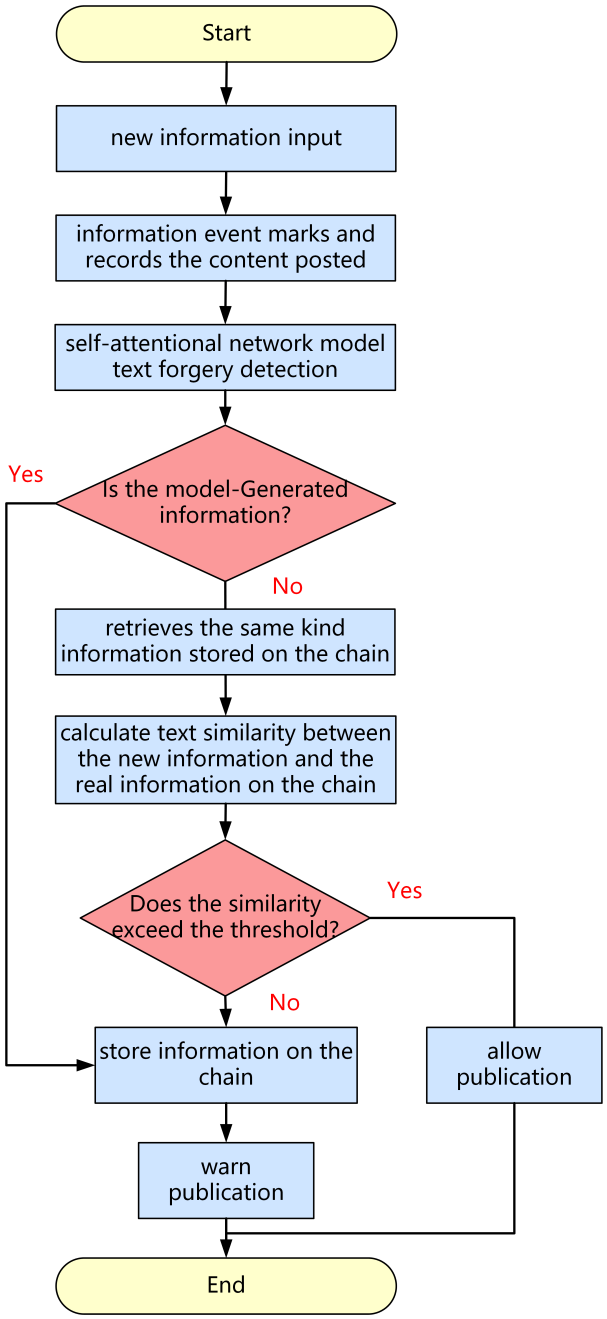


Figure 2. False information detection and control.

3.2.2. Model-Generated False Information Traceability

The false information traceability function mainly contains information event name marking, information similarity calculation, and blockchain traceability. The function process is shown in Figure 3. Firstly, the system uploads the false information content to the system for information event name marking. Secondly, the system queries whether false information with the same event name is already stored. If there is no storage on the chain, the uploaded information is registered and stored on the chain; otherwise, the text similarity detection algorithm is used to compare the similarity of the false information. If the information's similarity exceeds the threshold, the system retrieves the stored data on the chain to enable false information traceability and evidence extraction; otherwise, the false information is stored on the chain.

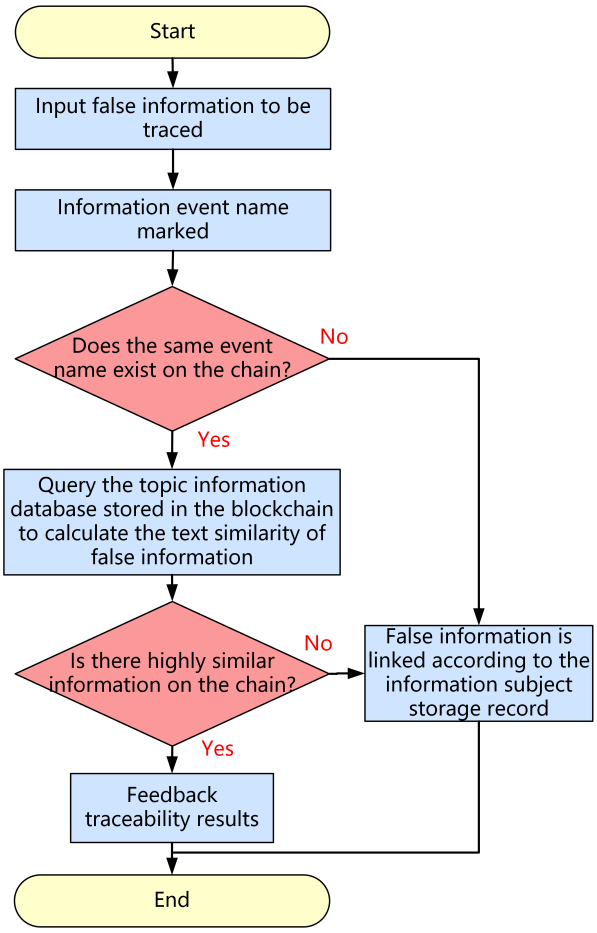


Figure 3. False information traceability.

4. Models

This section focuses on the false text detection model and text similarity detection model. These two key technologies play a key role in false information detection, which enables this system to recognize and stop false information in the process of information dissemination and restore the real information when necessary, ensuring the reliability and authenticity of information.

4.1. Self-Attention Network-Based False Text Discrimination

In this paper, we use the self-attention model to discriminate text, which includes text segmentation and mapping classification. The model structure is shown in Figure 4. The self-attention model uses the WordPiece lexer to segment the input text. The implementation principle involves splitting all the words in the glossary into a series of subwords, which prevents unrecorded words from being used and improves the model's generalization performance. After segmentation, all the subwords are converted into corresponding ID representations, which are input into the model to obtain the representation vector of the text. The process of text mapping classification consists of inputting the output of the self-attention model into a fully connected layer that maps the vectors to a binary classification probability distribution. The output layer typically uses a sigmoid function to map the resulting vectors to a range of values [0,1]. The type of prediction, i.e., model-generated content, is predicted for a probability greater than 0.5, and human-written content is predicted for a probability less than or equal to 0.5.

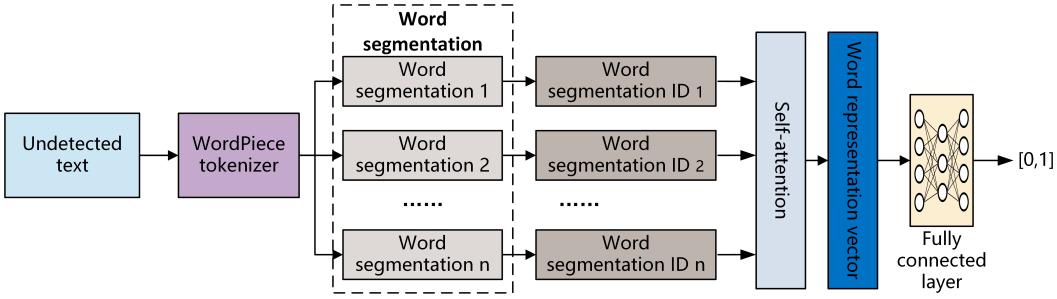


Figure 4. Self-attention network text detection model.

The process of text discrimination using the self-attention model in this project includes two processes: model training and model discrimination. The detailed algorithm is shown in Algorithm 1.

Algorithm 1: Self-attention model-based implementation of fine-tuned textual discrimination models

Input: training set X_{train}, test set X_{test}, label set Y_D
Output: text discrimination results $\hat{Y}_D$

Step 1: Data pre-processing
Map the training set X_{train} and test set X_test into word vectors to obtain X_{0train} and X_{0test}.
Step 2: Build the model
Define a model, set model hyperparameters, classification format, number of training rounds
Step 3: Train the model
Use K-FOLD to randomly divide the training set into 5 copies for training
while not reaching the preset number of training rounds **do**
  while not reaching the number of iterations **do**
    Set the number of samples for a single training batch $batch_size$ to use a small batch of dataset batch as model inputs
    Calculate the cross-entropy loss function, Classes denotes the number of network traffic classes
    if number of iterations %15==0 **then**
      learning rate = learning/10
    end if
    Updating Model Parameters Using the Adam Optimizer
    if acc>best acc for this iteration **then**
      Save the model for this iteration
      Update best acc.
    end if
  end while
  Setting up validation sets, validating models using validation sets and fine-tuning parameters
end while
Step 4: Save the model
Saving the model after parameter fine-tuning
Step 5: Test the model
Load the saved model and test the model with the test set
return text discriminant classification results $\hat{Y}_D$

The process of training the self-attention model includes two parts: data preparation and preprocessing, and model training and evaluation. Data preparation and preprocessing involve collecting the training dataset, including model-generated content and human-

written content, to ensure that the number of the two text types is balanced to avoid data distribution bias. Preprocessing of the collected text data involves the following:

- Text cleaning (removing special characters, punctuation marks, etc.);
- Word splitting (converting the text into sequences of words);
- Encoding (converting each word into its corresponding ID representation).

The processed dataset is used for model training. First, the preprocessed text data are fed into the model to obtain a representation vector of the text. Then, the vectors are mapped to a binary classification probability distribution by adding a fully connected layer and a sigmoid function to determine whether the text is model-generated or written by a human. The model parameters are tuned by optimizing the loss function (e.g., cross-entropy loss) to make predictions consistent with the actual labels. The trained model is evaluated using a validation dataset to assess the model's performance on the binary classification task.

4.2. Twin Network-Based Text Similarity Detection

After determining that a text is artificial, the system cannot determine whether the text is false information solely based on language logic. Therefore, to address the need for information traceability and to find the source of similar false information, many actual news texts released by credible, authoritative platforms obtained by crawlers are uploaded and stored. Then, using the statement similarity model, the manufactured text is matched against the currently detected text and the text on the chain. If the similarity exceeds the set threshold, it is determined that the descriptions of the two texts are the same, and the results of the text detection on the chain are returned. If no corresponding similar text is found, the current text content is considered false information. Otherwise, the text content is considered valid.

For this reason, this paper proposes a twin network-based text similarity calculation method, whose primary purpose is to compare two texts and evaluate their similarity to perform tasks such as text matching, semantic search, and text classification. Unlike the traditional text similarity calculation method, the utterance similarity model uses a twin network structure, i.e., a double parallel convolutional neural network, to encode the two texts and calculate the cosine similarity between the two utterances while preserving the textual semantics. This approach uses the structure of the Sentence-Bert utterance similarity model, as shown in Figure 5.

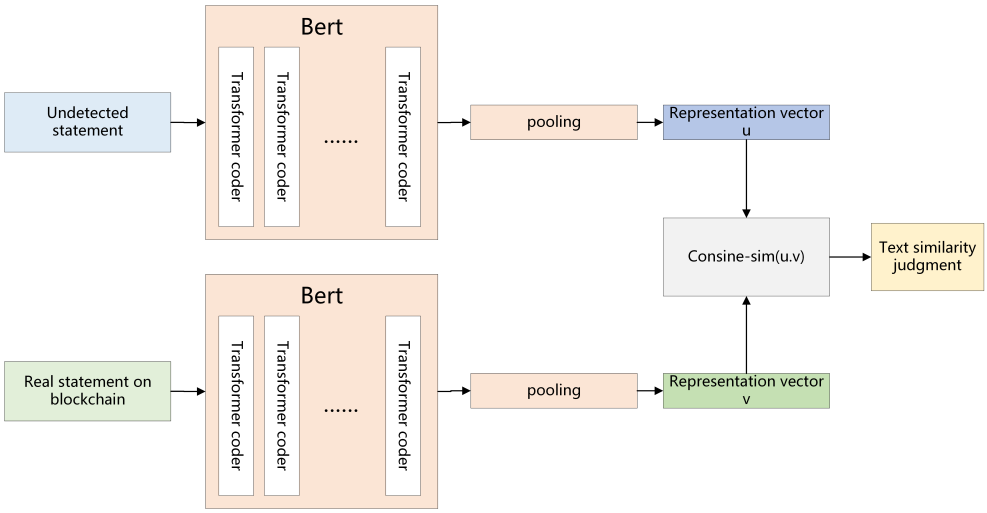


Figure 5. Text similarity detection model based on twin network.

We connect the sentence embeddings u and v with the element differences $|u - v|$ using trainable weights $w_t \in R^{3nk}$.

$$o = softmax(w_t(u, v, |u - v|)). \quad (1)$$

where n is the dimension of the sentence embedding and k is the number of labels.

Given an anchor sentence A, a positive sentence P, and a negative sentence N, the ternary loss function adjusts the network so that the distance between A and P is less than the distance between A and N. The loss function is a function of the mean square error. Mathematically, we minimize the following loss function:

$$max(||e_A - e_P|| - ||e_A - e_N|| + \theta, 0) \quad (2)$$

where e_A, e_P, and e_N are the embeddings of sentences a, n, and p. $||\cdot||$ is the Euclidean distance metric. θ ensures that e_P is at least θ closer to e_A than e_N.

Once model training is complete, generating a text similarity score between 0 and 1 is possible by inputting the detected text data and the text on the chain. We evaluate the impact of different thresholds on the model's performance through cross-validation to determine an appropriate similarity threshold of 0.6. If the threshold is exceeded, the text content is considered the same on both ends.

5. Experimental Evaluation

In order to verify the feasibility and performance index of the proposed blockchain-based model-generated false information detection and control system, this paper tests and evaluates the scheme through algorithm implementation and system simulation.

5.1. Environment

In this paper, we use a workstation with an Intel i9-13900HX processor, 32 GB of memory, an Nvidia GeForce RTX4060 GPU, and an Ubuntu 20.04 operating system VM using Hyper-V Manager. Four lightweight Fisco-BCOS blockchain nodes are deployed on the VM through Docker 2.9.1 to simulate the blockchain network environment of the system. In addition, this paper completes the development and testing of the proposed algorithm and system using Python 3.7, Pytorch 1.10.0, and Django 2.1.

5.2. Blockchain Architecture

When compared to the blockchain architectures of FISCO BCOS [27], Hyperledger Fabric [28], and Ethereum [29], FISCO BCOS demonstrates multiple advantages that make it more attractive for building decentralized applications. Table 1 shows the performance comparison of different blockchain architectures.

First, for the consensus mechanism, the Byzantine Fault-Tolerant (BFT) consensus algorithms adopted by FISCO BCOS, such as PBFT and rPBFT, provide a 1/3 fault-tolerance rate, which significantly enhances the system's stability and security. This better suits environments requiring high reliability and security compared to Hyperledger Fabric's non-Byzantine fault-tolerant consensus and Ethereum's proof of work (PoW).

Second, FISCO BCOS has a single-chain TPS of more than 20,000 regarding performance and scalability, far exceeding Hyperledger Fabric's 3400 TPS and Ethereum's PoW-constrained performance. In addition, FISCO BCOS theoretically supports an unlimited number of nodes, which provides greater scalability and flexibility for decentralized applications.

Third, regarding cross-chain solutions, FISCO BCOS supports isomorphic and heterogeneous cross-chain interactions through WeCross cross-chain routing. In contrast, Hyperledger Fabric and Ethereum's cross-chain solutions are relatively complex and rely on BaaS platforms, smart contracts, and DApps, which gives FISCO BCOS an advantage in cross-chain communication and data sharing.

Table 1. Performance comparison of FISCO BCOS, Hyperledger Fabric, and Ethereum.

Feature	Hyperledger Fabric	FISCO BCOS	Ethereum
Consensus Mechanism	Does not support Byzantine fault tolerance; maintains network security through consensus among organization members	Supports Byzantine fault tolerance; 1/3 fault tolerance rate; uses PBFT and rPBFT consensus algorithms	Proof of Work (PoW)
TPS (Transactions Per Second)	About 3400 (32-core CPU; 10 nodes)	Over 20,000 (single chain)	Limited by Proof of Work; limited scalability
Node Scalability	Weak node scalability; most deployed projects have single-digit nodes	Theoretically, the number of nodes is unlimited	Limited by Proof of Work; limited scalability
Cross-chain Solution	Homogeneous cross-chain depends on BaaS platform	WeCross cross-chain routing; supports homogeneous and heterogeneous cross-chain interactions	Complex; depends on smart contracts and DApps
Deployment Support	Hyperledger Explorer blockchain browser; BaaS platform	WeBASE, WeIdentity, WeEvent, WeCross, and other solutions	Community and third-party tools, such as Truffle and Mist
Storage	Block storage is saved in file format; supports LevelDB and CouchDB storage	Supports multiple storage solutions; optimized storage performance	Stored on each node; performance is limited when the data volume is large
Smart Contracts	ChainCode; Docker deployment; supports multiple languages	Solidity contracts; precompiled contracts; supports parallel computing	EVM execution; smart contracts are widespread; supports multiple languages
Performance Optimization	Reduces key conflicts, reduces stub reads, and writes to the ledger	Transaction broadcast strategy optimization; load balancing; callback decoupling; signature verification deduplication	Community explores scaling solutions such as sharding technology

Fourth, regarding deployment support, FISCO BCOS provides solutions such as WeBASE, WeIdentity, WeEvent, etc., which provide enterprises with comprehensive blockchain application deployment and management tools. In contrast, Hyperledger Fabric and Ethereum rely heavily on community and third-party tools, which may increase the complexity and cost of deployment and management for enterprises.

Fifth, in terms of storage, FISCO BCOS supports multiple storage solutions and optimizes storage performance. In contrast, Ethereum's storage needs to be performed on each node, which may affect performance when data volumes are large. It gives FISCO BCOS an advantage in handling large-scale data.

Sixth, FISCO BCOS supports Solidity contracts, precompiled contracts, and parallel computing for smart contracts, making their execution more efficient. Although Hyperledger Fabric and Ethereum also support smart contracts, FISCO BCOS may be better regarding parallel processing and performance optimization.

Finally, regarding performance optimization, FISCO BCOS significantly improves the system's overall performance through transaction broadcasting policy optimization, load balancing, callback stripping, and signature verification deduplication. These optimizations make FISCO BCOS more stable and efficient in handling highly concurrent transactions.

In summary, this paper selects FISCO BCOS as the blockchain architecture for the prototype system because it shows significant advantages in consensus mechanisms, performance and scalability, cross-chain solutions, deployment support, storage, smart contracts, and performance optimization.

5.3. Dataset

This paper uses FakeNewsNet [30] for the information detection dataset. FakeNewsNet contains two sub-datasets from two fact-checking websites: GossipCop and PolitiFact. Both sub-datasets contain news content, dissemination information on social media, and spatio-temporal information, as well as true–false labels, labeled by professional journalists and experts. In addition, the FakeNewsNet dataset contains not only textual features

but also user behavior, network structure, and sentiment features on social media, which provide rich information for online false information detection. In this paper, we divide the data for training, validation, and testing in the ratio of 7:2:1 for both sub-datasets.

5.4. Indicators

In this paper, we use *accuracy* and *F*1 *score* for the binary classification model performance evaluation, which are calculated as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F1 - score = \frac{2 * precsion * recall}{precision + recall} \tag{4}$$

where true positives (TP) represent the number of predicted false information instances that are actually false, true negatives (TN) represent the number of predicted true information instances that are actually true, false positives (FP) represent the number of predicted false information instances that are actually true, and false negatives (FN) represent the number of predicted true information instances that are actually false. *precision* denotes the model detection precision rate and *recall* denotes the model detection recall rate. The equations for these metrics are as follows:

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

In addition, in terms of the blockchain system's performance, the stability of the system is tested through task processing delay and data throughput.

5.5. Results

In this paper, the bias of the experimental results was reduced by calculating the average of the results from five repetitions of the experiments in order to compare and analyze the performance of the model and the system.

5.5.1. False Information Detection Results

Using existing studies [31–34], the measurement accuracy and F1 score on two sub-datasets of GossipCop and PolitiFact were compared experimentally, and the comparison results are shown in Table 2.

Table 2. Comparison of detection performance.

	GossipCop		**PolitiFact**	
	Accuracy	**F1 Score**	**Accuracy**	**F1 Score**
[31]	0.8472	0.8532	0.8216	0.8251
[32]	0.8613	0.8564	0.8401	0.8334
[33]	0.7873	0.7654	0.8245	0.8245
[34]	0.8601	0.7438	0.9019	0.7691
Ours	0.8546	0.8615	0.8360	0.8476

Meanwhile, the test analysis found that the accuracy of the proposed algorithm was slightly lower than that in [32,34] on the GossipCop dataset and lower than that in [34] on the PolitiFact dataset. However, compared with [31,33], it demonstrated a clear advantage.

In addition, the proposed algorithm achieved the highest F1 score on the two datasets, which indicates that the proposed scheme can better balance precision and recall, and the model has higher performance.

In summary, the proposed algorithm's overall performance on the GossipCop and PolitiFact datasets was more stable. It maintained a certain degree of accuracy and achieved a high F1 score, which indicates that the algorithm has a strong generalization ability and adaptability.

5.5.2. Quality-of-Service Results

In this paper, based on the created blockchain network architecture, the Hyperledger Caliper tool was used to test the latency and throughput of the system's information storage addition and information traceability query functions under different task volumes. The test results are shown in Figures 6 and 7, respectively.

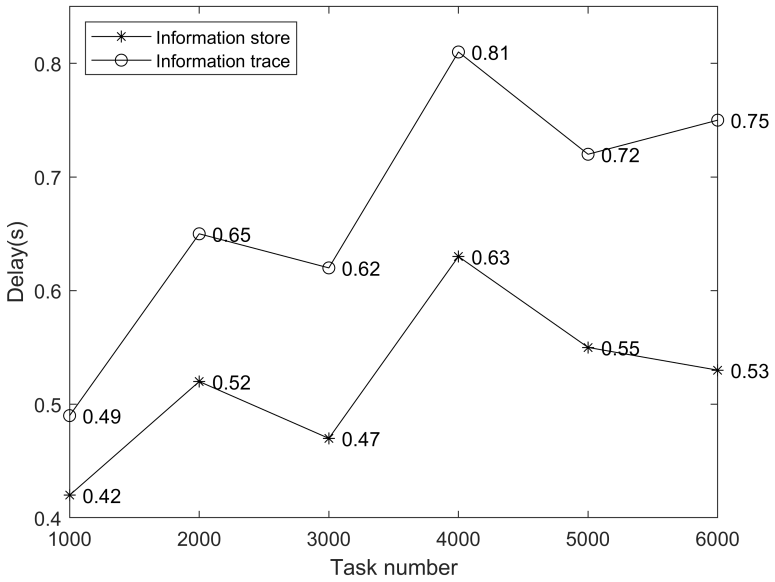


Figure 6. System delay.

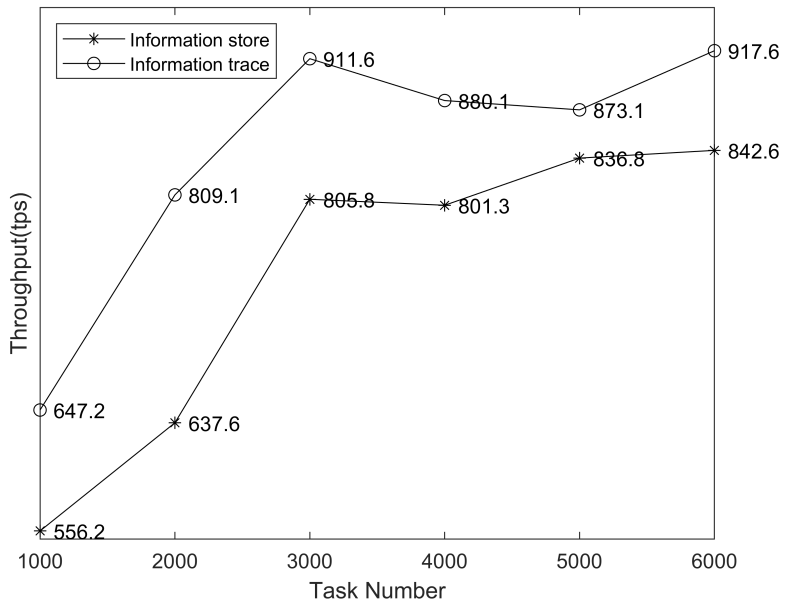


Figure 7. System throughout.

In Figures 6 and 7, we can see that for the same task volume, information traceability requires more delay and throughput compared to information storage. This is due to the need to retrieve the index query of the information during information traceability,

increasing the system overhead. With the continuous growth of the task volume, although task processing latency and throughput show an upward trend, they are still within a reasonable range, indicating the feasibility and stability of the system proposed in this paper. In addition, information throughput tends to level off from 3000 tasks. This paper suggests that this phenomenon occurs because the number of system tasks reaches the overall performance bottleneck of this block simulation network.

The false information detection and blockchain task processing tests showed that the blockchain-based false information detection and control system scheme proposed in this paper has performance advantages and application value.

6. Conclusions and Future Work

In this paper, we propose a blockchain-based false information detection and control system. Compared with existing research, the system proposed in this paper provides early warning before the release of false information and information traceability after its release. This addresses the shortcomings of existing research on early warning before the release of false information and accountability after its release. In addition, this paper also innovatively combines a false information detection method based on a self-attention network with text similarity detection based on a twin network. This method guarantees the accuracy of model-generated false information detection and similarity comparison, which effectively supports the application function of the system. The false information detection and blockchain task processing experiments prove that the model and system proposed in this paper have stable detection performance, and the prototype system is feasible for practical application.

However, our work still has some limitations. The accuracy of the proposed method needs improvement through hyperparameter tuning, using additional datasets, and optimizing the model's architecture. The blockchain-based model-generated false information detection and control system designed in this paper considers all model-generated texts as false. This implementation method may incorrectly classify legitimate model-generated information as false, which reduces detection accuracy. In addition, model-generated false information is multidimensional, encompassing not only text information but also voice, image, and video. Therefore, our future work will focus on improving the model's accuracy and developing determination mechanisms for false information across various media.

Author Contributions: C.L. developed the idea, conducted the research, and wrote the manuscript. Y.X. and B.H. conducted the analyses. Z.S. verified and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (No. 62272239, No. 62302237), Guizhou Provincial Key Technology R&D Program (No. [2023]272), the Jiangsu Agriculture Science and Technology Innovation Fund (No. CX(22)1007), the Natural Science Research Start-up Foundation of Recruiting Talents of Nanjing University of Posts and Telecommunications (No. NY222029), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 22KJB520027), and the Postgraduate Research and Innovation Plan of Jiangsu Province (No. KYCX20_0761).

Data Availability Statement: The original data presented in the study are openly available in FakeNewsNet at https://github.com/KaiDMML/FakeNewsNet (accessed on 14 June 2024).

Acknowledgments: We wish to thank all dataset providers. We also wish to thank all colleagues, reviewers, and editors who provided valuable suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. Chai, Y.; Liu, Y.; Li, W.; Zhu, B.; Liu, H.; Jiang, Y. An interpretable wide and deep model for online disinformation detection. *Expert Syst. Appl.* **2024**, *237*, 121588. [CrossRef]
2. Kumar, V.; Sinha, D. Synthetic attack data generation model applying generative adversarial network for intrusion detection. *Comput. Secur.* **2023**, *125*, 103054. [CrossRef]
3. Google. Google JIGSAW. Available online: https://jigsaw.google.com/ (accessed on 10 July 2024).
4. Murayama, T.; Wakamiya, S.; Aramaki, E.; Kobayashi, R. Modeling the spread of fake news on Twitter. *PLoS ONE* **2021**, *16*, e0250419. [CrossRef] [PubMed]
5. Meta. Here's How We're Using AI to Help Detect Misinformation. Available online: https://ai.meta.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/ (accessed on 10 July 2024).
6. Liu, J.; Ke, J.; Liu, J.; Xie, X.; Tian, E. Outlier-Resistant Non-Fragile Control of Nonlinear Networked Systems Under DoS Attacks and Multi-Variable Event-Triggered SC Protocol. *IEEE Trans. Inf. Forensics Secur.* **2024**, *19*, 2609–2622. [CrossRef]
7. Phan, H.T.; Nguyen, N.T.; Hwang, D. Fake news detection: A survey of graph neural network methods. *Appl. Soft Comput.* **2023**, *139*, 110235. [CrossRef] [PubMed]
8. Li, Q.; Gao, M.; Zhang, G.; Zhai, W.; Chen, J.; Jeon, G. Towards Multimodal Disinformation Detection by Vision-language Knowledge Interaction. *Inf. Fusion* **2024**, *102*, 102037. [CrossRef]
9. Zhuang, P.; Zamir, T.; Liang, H. Blockchain for cybersecurity in smart grid: A comprehensive survey. *IEEE Trans. Ind. Informatics* **2020**, *17*, 3–19. [CrossRef]
10. Tian, Z.; Li, M.; Qiu, M.; Sun, Y.; Su, S. Block-DEF: A secure digital evidence framework using blockchain. *Inf. Sci.* **2019**, *491*, 151–165. [CrossRef]
11. Sunny, J.; Undralla, N.; Pillai, V.M. Supply chain transparency through blockchain-based traceability: An overview with demonstration. *Comput. Ind. Eng.* **2020**, *150*, 106895. [CrossRef]
12. Ryu, J.H.; Sharma, P.K.; Jo, J.H.; Park, J.H. A blockchain-based decentralized efficient investigation framework for IoT digital forensics. *J. Supercomput.* **2019**, *75*, 4372–4387. [CrossRef]
13. Malik, A.; Sharma, A.K. Blockchain-based digital chain of custody multimedia evidence preservation framework for internet-of-things. *J. Inf. Secur. Appl.* **2023**, *77*, 103579.
14. Dwivedi, A.D.; Singh, R.; Dhall, S.; Srivastava, G.; Pal, S.K. Tracing the source of fake news using a scalable blockchain distributed network. In Proceedings of the 2020 IEEE 17th International Conference on Mobile ad Hoc and Sensor Systems (MASS), Delhi, India, 10–13 December 2020; pp. 38–43.
15. Chen, Q.; Srivastava, G.; Parizi, R.M.; Aloqaily, M.; Al Ridhawi, I. An incentive-aware blockchain-based solution for internet of fake media things. *Inf. Process. Manag.* **2020**, *57*, 102370. [CrossRef]
16. Shahbazi, Z.; Byun, Y.C. Fake Media Detection Based on Natural Language Processing and Blockchain Approaches. *IEEE Access* **2021**, *9*, 128442–128453. [CrossRef]
17. Shrivastava, G.; Kumar, P.; Ojha, R.P.; Srivastava, P.K.; Mohan, S.; Srivastava, G. Defensive modeling of fake news through online social networks. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 1159–1167. [CrossRef]
18. Verma, P.K.; Agrawal, P.; Amorim, I.; Prodan, R. WELFake: Word embedding over linguistic features for fake news detection. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 881–893. [CrossRef]
19. Guo, Z.; Zhang, Q.; Ding, F.; Zhu, X.; Yu, K. A Novel Fake News Detection Model for Context of Mixed Languages Through Multiscale Transformer. *IEEE Trans. Comput. Soc. Syst.* **2023**, 1–11. [CrossRef]
20. Balshetwar, S.V.; Rs, A. Fake news detection in social media based on sentiment analysis using classifier techniques. *Multimed. Tools Appl.* **2023**, *82*, 35781–35811. [CrossRef] [PubMed]
21. Raza, S.; Ding, C. Fake news detection based on news content and social contexts: A transformer-based approach. *Int. J. Data Sci. Anal.* **2022**, *13*, 335–362. [CrossRef]
22. Chen, C.C.; Du, Y.; Peter, R.; Golab, W. An implementation of fake news prevention by blockchain and entropy-based incentive mechanism. *Soc. Netw. Anal. Min.* **2022**, *12*, 114. [CrossRef]
23. Sengupta, E.; Nagpal, R.; Mehrotra, D.; Srivastava, G. ProBlock: A novel approach for fake news detection. *Clust. Comput.* **2021**, *24*, 3779–3795. [CrossRef]
24. Duvvuri, V.N.D.; Irfan, M.; Surakarapu, V.; Kota, B.; Pyla, J.; Ponnapalli, M. Detecting Fake News Using Blockchain and Deep Learning. In Proceedings of the 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 23–25 August 2023; IEEE: New York, NY, USA, 2023; pp. 1308–1311.

25. Dhall, S.; Dwivedi, A.D.; Pal, S.K.; Srivastava, G. Blockchain-based framework for reducing fake or vicious news spread on social media/messaging platforms. *Trans. Asian -Low-Resour. Lang. Inf. Process.* **2021**, *21*, 1–33. [CrossRef]
26. Xiao, Y.; Liu, Y.; Li, T. Edge computing and blockchain for quick fake news detection in IoV. *Sensors* **2020**, *20*, 4360. [CrossRef] [PubMed]
27. Li, H.; Chen, Y.; Shi, X.; Bai, X.; Mo, N.; Li, W.; Guo, R.; Wang, Z.; Sun, Y. FISCO-BCOS: An Enterprise-grade Permissioned Blockchain System with High-performance. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '23, Denver, CO, USA, 11–17 November 2023. [CrossRef]
28. Androulaki, E.; Barger, A.; Bortnikov, V.; Cachin, C.; Christidis, K.; De Caro, A.; Enyeart, D.; Ferris, C.; Laventman, G.; Manevich, Y.; et al. Hyperledger fabric: A distributed operating system for permissioned blockchains. In Proceedings of the Thirteenth EuroSys Conference, EuroSys '18, Porto, Portugal, 23–26 April 2018. [CrossRef]
29. Buterin, V. Ethereum: Platform review. *Oppor. Challenges Priv. Consort. Blockchains* **2016**, *45*, 1–45.
30. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. *ACM Sigkdd Explor. Newsl.* **2017**, *19*, 22–36. [CrossRef]
31. Lee, J.; Toutanova, K. Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
32. Yao, L.; Mao, C.; Luo, Y. Graph convolutional networks for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu , HI, USA, 27 January–1 February 2019; Volume 33, pp. 7370–7377.
33. Ni, S.; Li, J.; Kao, H.Y. True or False: Does the Deep Learning Model Learn to Detect Rumors? In Proceedings of the 2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taichung, Taiwan, 18–20 November 2021; pp. 119–124. [CrossRef]
34. Han, Y.; Karunasekera, S.; Leckie, C. Graph neural networks with continual learning for fake news detection from social media. *arXiv* **2020**, arXiv:2007.03316.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.