applied sciences

MDPI

Article

RumorLLM: A Rumor Large Language Model-Based Fake-News-Detection Data-Augmentation Approach

Jianqiao Lai, Xinran Yang, Wenyue Luo, Linjiang Zhou, Langchen Li, Yongqi Wang and Xiaochuan Shi *

School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China
* Correspondence: shixiaochuan@whu.edu.cn

Abstract: With the rapid development of the Internet and social media, false information, rumors, and misleading content have become pervasive, posing significant threats to public opinion and social stability, and even causing serious societal harm. This paper introduces a novel solution to address the challenges of fake news detection, presenting the "Rumor Large Language Models" (RumorLLM), a large language model finetuned with rumor writing styles and content. The key contributions include the development of RumorLLM and a data-augmentation method for small categories, effectively mitigating the issue of category imbalance in real-world fake-news datasets. Experimental results on the BuzzFeed and PolitiFact datasets demonstrate the superiority of the proposed model over baseline methods, particularly in F1 score and AUC-ROC. The model's robust performance highlights its effectiveness in handling imbalanced datasets and provides a promising solution to the pressing issue of false-information proliferation.

Keywords: fake-news detection; large language models; rumor generation; category imbalance; data augmentation

Citation: Lai, J.; Yang, X.; Luo, W.; Zhou, L.; Li, L.; Wang, Y.; Shi, X. RumorLLM: A Rumor Large Language Model-Based Fake-News-Detection Data-Augmentation Approach. *Appl. Sci.* **2024**, *14*, 3532. https://doi.org/10.3390/app14083532

Academic Editors: Dionisios Sotiropoulos and Douglas O'Shaughnessy

Received: 24 February 2024
Revised: 29 March 2024
Accepted: 8 April 2024
Published: 22 April 2024

1. Introduction

With the rapid development of the Internet and social media, information is now being generated and disseminated at an unprecedented rate, while the cost of dissemination has fallen significantly [1,2]. Unfortunately, this has also led to the proliferation of false information, rumors, and misleading content [3]. These deceptive elements can mislead the public [4], disrupt social stability, and even lead to serious consequences such as human casualties and national disruption [5,6]. Therefore, it is crucial to address the issue of fake news and develop effective methods for its detection [7,8].

There are several challenges associated with the detection of fake news. First, real-world fake-news datasets typically contain fewer samples of fake news compared to real news [9], resulting in an imbalanced distribution. This imbalance can bias the performance of the classifier [10], reducing its accuracy in predicting less-common categories. In addition, the larger number of real-news samples can cause the classifier to incorrectly predict new samples as real news when they are actually fake news. This misclassification can lead to significant social harm [5,11].

To address these challenges, this paper proposes the construction of a rumor-generating large language model called the "Rumor Large Language Model" (RumorLLM). This model is finetuned using rumor-specific writing styles, content, and general semantic information. By exploiting the exceptional capabilities of large language models in natural-language processing tasks, such as capturing lexical relations, understanding context, performing semantic reasoning, and exhibiting strong generalization ability, we aim to improve the accuracy and efficiency of fake-news detection. In this approach, we use the large language model for data augmentation, specifically for the task of fake-news detection.

To summarize the innovations and contributions of this paper:

1. We construct a rumor-generating large language model, named "Rumor Large Language Models" (RumorLLM), by employing a hybrid finetuning approach that incorporates the writing style and content of rumors. This model fills the existing gap in large language models specifically tailored for fake news detection.
2. We propose a method based on RumorLLM and prompt engineering to diversify and enhance the small categories of samples. This approach enhances the model's ability to discriminate complex rumors generated by artificial intelligence.
3. We ensemble RumorLLM with state-of-the-art classification models and validate the effectiveness of our methods using real datasets.

The remainder of this paper is structured as follows: Section 2 provides an overview of related work on fake news detection, with a particular emphasis on addressing data imbalance. Section 3 presents our proposed approach in detail. Section 4 describes the dataset used and presents an analysis of baselines, along with experimental results. Finally, Section 5 concludes the paper.

2. Related Work

Over the past decade, there have been significant efforts to use machine learning methods to detect fake news. Benchmark models such as support vector machines (SVMs) and stochastic gradient descent (SGD) [12], ordinary Bayesian classifiers [13], and decision tree algorithms [14] have been widely used in the field. However, earlier studies on fake news detection usually relied on the manual extraction of relevant textual information and the manual creation of features, which were then fed into the aforementioned machine learning models for classification. For example, Castillo et al. in 2011 [15] manually extracted features such as punctuation and word count as indicators of news authenticity.

In the field of fake news detection, supervised learning is currently the mainstream method. It is mainly divided into two main categories: methods based on traditional machine-learning methods [16] and methods based on deep-learning methods [17,18]. In previous studies, text and user information were mainly extracted using statistical machine learning or neural networks to extract textual features [19]. For example, linguistic features were manually selected, and only textual information was used for classification [16]. With the advent of deep learning, recurrent neural networks (RNNs) were introduced to capture hidden representations in text features [17]. Subsequently, several studies used Convolutional Neural Networks (CNNs) for fake news detection by mapping each post of a news event to a vector space and using CNNs to extract textual features from the resulting embedding matrix, which are then fed into a classifier for final classification [20]. Another approach proposes a Graph Convolutional Network (GCN) model, which represents news articles as a graph with sentences as nodes and similarities between sentences as edges, transforming fake news detection into a graph classification problem [21]. In addition, Alzanin et al. in 2019 [22] also used semi-supervised and unsupervised methods to detect fake news in social media.

With the development of deep-learning technology in recent years, neural network modeling has become a mainstream method for detecting fake news. Researchers have used models such as CNN and RNN to process and detect fake news. Although these studies have achieved promising results, most of them have mainly focused on text features [23] and ignored the potential benefits of combining image features [24]. However, in the social media domain, news articles accompanied by images tend to be more widely disseminated due to their visual appeal and the different viewpoints they convey [25]. In addition, images often contain richer semantics [26]. Therefore, many researchers have started to focus on the role of images in fake news detection and have proposed multimodal detection models [27,28]. For example, Jin et al. in 2017 [29] pioneered the use of an RNN-based fake-news-detection model that uses both textual and image information to determine the authenticity of news. In addition, some researchers have introduced mechanisms such as attention [30] and text-image consistency to achieve better results using text and image information more effectively [31,32].

However, none of the above studies have taken into account the common problem of fake news in the real world—that is, the problem of an unbalanced distribution of fake-news samples in the real world—and the common practice is to resample this category of data from a few samples using three oversampling strategies: Random oversampling; generating synthetic samples from a few categories using k-nearest neighbor methods; oversampling by generating their distribution based on the distribution of the few categories of synthetic samples, oversampling by generating the distribution of the few categories of data samples based on their distributions [33], or even resampling the latent spatial representations mapped by deep learning to balance the dataset by resampling the hidden vectors using a variety of resampling techniques including oversampling, under-sampling and hybrid sampling [11]. However, resampling methods are a single re-use of a particular piece of data, which can introduce noise or cause overfitting problems and do not always improve the performance of the model [34]. Another solution to category imbalanced samples is to use some conventional text enhancement techniques, such as translating into other languages, randomly inserting some new words or randomly deleting some words, or randomly changing the order of some words, etc. [10] . However, it has been pointed out by some scholars [9] that this method probably performs poorly because of the high dimensionality of the bilingual space. Random insertion and deletion to increase the amount of data does not always have the effect of improving the prediction performance [34] because not all the newly data are favorable to the train.

In summary, while previous research has made significant progress in fake news detection using machine learning and deep-learning methods, the issue of the imbalanced distribution of fake-news samples has not been adequately addressed. This paper proposes a novel approach that leverages a rumor-generating large language model to tackle this problem and enhance the accuracy of fake news detection. The proposed approach builds upon existing research by incorporating advanced language modeling techniques and addressing the limitations of traditional resampling and text enhancement methods.

3. Methodology

To address the problems mentioned in the above sections, this paper proposes to construct a rumor-generating large language model (hereafter referred to as "RumorLLM") by finetuning large language models (LLMs), incorporating rumor writing styles and contents as well as the general semantic information of LLMs. A rumor-generating large language model called "RumorLLM" (hereafter referred to as RumorLLM) was constructed. Large language models have demonstrated remarkable capabilities in natural-language processing tasks in recent years. The success of these models has contributed prominently to many research areas covering a variety of topics such as architectural innovations in the underlying neural networks, context length improvement, model alignment, training datasets, benchmarking, efficiency, etc. [35]. Large language models especially excel in natural-language processing tasks because these large language models are able to capture complex relationships between words, better contextual comprehension, more complex semantic reasoning, more excellent generalization, and generate more coherent and consistent linguistic text to improve the accuracy and efficiency of NLP tasks [36], LLMs demonstrate strong language comprehension and language generation capabilities So we use LLM to do data augmentation for the task of fake news detection. The model RumorLLM utilizes the excellent semantic understanding, analysis, and generation capabilities of common large language models (such as the ChatGLM series, LLaMA series, etc.) and also takes into account rumor-specific writing styles and content characteristics (using local-parameter finetuning based on LoRA [37], P tuningV2 [38], etc.), and its main structure is formulated as shown in Figure 1.

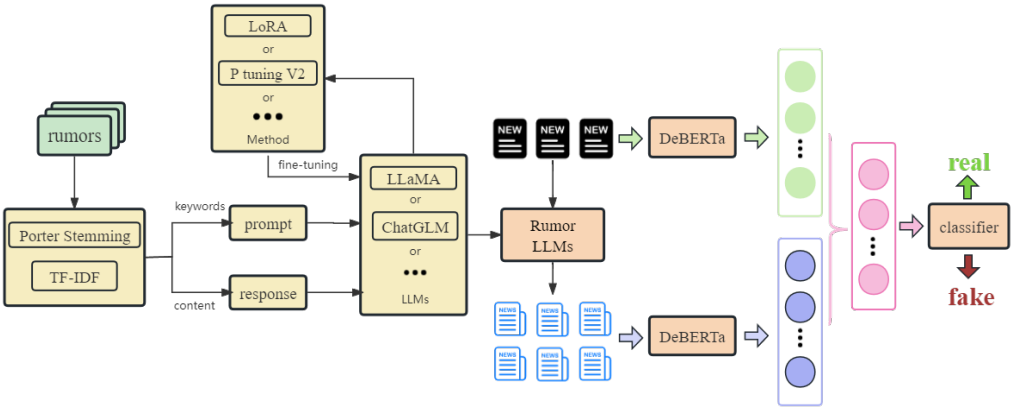

Figure 1. The methodology flow proposed in this paper.

In this way, we can generate text that better matches the characteristics of rumors, thus improving the accuracy and efficiency of rumor detection. This approach not only provides a new method for rumor generation and detection but also provides new perspectives and tools for us to understand and respond to the spread of false information. Meanwhile, to address the problem of category imbalance, this paper proposes a diverse small-category sample enhancement method based on "RumorLLM". This method generates diversified small-category samples through RumorLLM and prompts engineering to increase the number of small-category samples, which can effectively improve the category imbalance problem of the dataset and enhance the prediction ability of the rumor-detection model. In addition, since the new samples are generated by RumorLLM, a large language model, constructing reasonable classifiers for this part of the data can effectively improve the model's ability to discriminate those complex rumors generated by artificial intelligence.

3.1. Finetune

To enhance the performance of our target rumor generation language model (RumorLLM), we utilize a set of preprocessing and finetuning procedures outlined in Algorithm 1. In this section, we detail the process of constructing the corresponding prompt and response as a finetuned dataset using the Porter Stemming algorithm and the TF-IDF algorithm.

The Porter Stemming algorithm is a classical algorithm for stemming extraction that aims to reduce words to their original stemmed form. By removing affixes and word endings, we can obtain more concise key information, which helps to reduce the complexity of the vocabulary. The steps of the Porter Stemming algorithm include regular substitution, applying a series of regular substitution operations to reduce the affixes and endings of words; conditional rule application, applying rules based on specific conditions to ensure that the word is processed correctly; and suffix processing: processing the suffixes of words to eliminate redundant endings.

The regular substitution of the Porter Stemming algorithm can be expressed by the following equation:

$$NewWords = ApplyRules(OldWords) \quad (1)$$

For ApplyRules, in conjunction with the Porter Stemming algorithm, this paper sets the following rules:

1. Common suffix deletion: delete common suffixes at the end of words, such as '-ed' and '-ing';
2. Form conversion: such as converting the plural form of a noun to the singular form;
3. Noun and verb suffix deletion: delete specific suffixes at the end of a noun or a verb, such as '-ness' and '-ize'.
4. Deletion of terminative suffixes: delete terminative suffixes at the end of a word if certain conditions are met, such as '-ant' and '-ent'.

Algorithm 1 Fine-tuning Process with Porter Stemming and TF-IDF

Input: Raw news text segmentation data $OldWords$, $OriginalPrompt$, vocabulary t, current document d, the entire set of documents D , Base large language model $BaseLLM$
Output: RumorLLM

1: $NewWords \leftarrow$ Apply a series of systematic replacement operations to $OldWords$.
2: **Porter Stemming Formula Representation:**

$$NewWords = ApplyRules(OldWords)$$

3: Calculate Term Frequency (TF): Compute the frequency of each word in the current document.
4: Calculate Inverse Document Frequency (IDF): Compute the inverse document frequency of each word in the entire document collection.
5: Calculate TF-IDF Value: Multiply term frequency and inverse document frequency to obtain the final TF-IDF value.
6: **TF-IDF Formula Representation:**

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D)$$

7: Based on the $NewWords$ and TF-IDF values, the keywords are reconstructed from largest to smallest according to the TF-IDF values to create prompts and responses for RumorLLM training.
8: **prompt_finetune Formula:**

$$prompt_{finetune} = Reconstruct(OriginalPrompt, NewWords, \text{TF-IDF})$$

9: Using the raw rumor text as the *response*, the final finetuning of the Base large language model yields the RumorLLM

$$RumorLLM = finetune(BaseLLM, prompt_{finetune}, response)$$

The TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is used to measure the importance of a word in the entire document collection. Words with high TF-IDF values usually indicate that the word is significantly informative in the current document.

The steps of the TF-IDF algorithm include calculating the word frequency (TF), calculating the frequency of each word in the current document, calculating the inverse document frequency (IDF): calculating the frequency of the inverse document for each word in the whole set of documents; calculating the TF-IDF value: multiplying the word frequency and the inverse document frequency, to obtain the final TF-IDF value.

The calculation of the TF-IDF algorithm can be represented by the following equation:

$$\text{TF-IDF}(t, d, D) = TF(t, d) \times IDF(t, D) \tag{2}$$

where t is the vocabulary, d is the current document, and D is the entire set of documents. The TF formula calculates the frequency of a term t in a given document d, and the IDF formula evaluates the rarity of a term t in the entire document collection D.

After Porter Stemming and TF-IDF processing, we integrate the keywords into a list of keywords that will be used to reconstruct the prompt and response. The prompt and response are reconstructed using the list of keywords to better reflect the rumor writing style and content characteristics, such as "Write a rumor with xxx as the keywords...". The formula used to finetune the prompt is represented as follows:

$$prompt_{finetune} = Reconstruct(OriginalPrompt, NewWords, \text{TF-IDF}) \tag{3}$$

Finally, we use the method of p-tuning V2 to finetune the target rumor generation language model to improve its performance and generation quality.

Through the above series of steps, we successfully apply Porter Stemming and TF-IDF algorithms to the finetuning process to construct more refined and information-rich rumor generation language models.

3.2. Data Augmentation

To reconstruct the rumor, we first utilize the word stems obtained through the Porter Stemming algorithm. These stems serve as the basis for preserving the core structure and meaning of the original rumor. By replacing specific words in the rumor with their corresponding stems, we ensure that the reconstructed rumor maintains semantic coherence while introducing variations.

Next, we incorporate the extracted keywords into the reconstructed rumor. These keywords provide crucial information that helps shape the content of the augmented rumor. By strategically placing the keywords within the reconstructed text, we emphasize their relevance and ensure that the generated rumor aligns with the original rumor's topic and focus.

Additionally, we leverage the finetuned RumorLLM to make appropriate expansions and generate new content for the reconstructed rumor. The language model incorporates its knowledge of rumor characteristics, writing style, and the prompt to generate plausible and engaging text. By combining the prompt, the original rumor's stem and keyword information, and the language model's capabilities, we create an augmented rumor that embodies the style and content characteristics of rumors while introducing new information.

The process of reconstructing and expanding the rumor can be represented by the following formula:

$$NewRumor = RumorLLM(Reconstruct(NewWords, \text{TF-IDF}), Prompt) \tag{4}$$

In this formula, RumorLLM is applied to the reconstructed rumor generated by combining the original rumor's stem and keyword information with the prompt. RumorLLM leverages its training and finetuning to generate a new rumor that aligns with the desired writing style and content characteristics.

By incorporating RumorLLM in the rumor generation process, we can transform the original rumor into a new rumor that reflects the language model's understanding of rumors and its ability to generate plausible and engaging text. This approach allows for the refinement and expansion of the rumor while maintaining its essence and characteristics, therefore creating a diverse training dataset for the RumorLLM.

3.3. News Representation

DeBERTa (Decoding-enhanced BERT with disentangled attention) [39] demonstrates significant advantages in solving the category imbalance problem in the fake-news-detection task. Its unique representation learning capability can better capture complex relationships in text through a decoupled attention mechanism. In the fake-news-detection task, DeBERTa improves the understanding of rumor-specific writing styles and content features by introducing location awareness and modeling long-distance dependencies. This is crucial for effectively distinguishing between different categories of text in the presence of category imbalance. Among them, DeBERTa's attention-mechanism formula can be expressed as:

$$W_{final} = block - diag(W_{in}, W_{out}, ..., W_{out}) \in \mathbb{R}^{n \times n} \tag{5}$$

$$Representation = news_i^T \cdot W_{final} \tag{6}$$

Equation (5) presents the computation of the final attention–weight matrix, denoted as W_{final}, in the context of utilizing the DisentangledAttention mechanism. This matrix is constructed by arranging multiple attention–weight matrices in a block-diagonal manner. Specifically, W_{in} represents the internal attention–weight matrix, while W_{out} represents

the external attention–weight matrix. The term "block-diag" signifies the arrangement of these attention–weight matrices, with W_{in} positioned on the diagonal and W_{out} located off the diagonal. The attention–weight matrix has a dimension of $\mathbb{R}^{n\times n}$, where n denotes the length of the input sequence.

Equation (6) defines the representation of the text, denoted as *Representation*, which is obtained by transposing the input news text $news_i$ ($news_i^T$) and multiplying it with the attention–weight matrix W_{final}. This operation allows for a weighted sum of the individual word representations based on the attention weights, therefore yielding the overall representation of the text.

To summarize, Equation (5) outlines the construction of the final attention–weight matrix, while Equation (6) elucidates how the attention–weight matrix is employed to perform a weighted sum of the input text, therefore generating the representation of the text. These equations constitute crucial computational steps within the DisentangledAttention mechanism, facilitating the capture of internal and external dependencies within the input text and generating corresponding text representations.

DeBERTa's Disentangled Attention makes the model more adaptable to the key information of different classes of text by decomposing this mechanism. In addition, DeBERTa's location-aware and decoupled design helps the model to better understand the contextual relationships in long texts, which improves its performance in fake news detection. In coping with the category imbalance problem, DeBERTa also effectively integrates the original data and the text generated by RumorLLM through its fusion approach of efficient parameter utilization and attention mechanism. This fusion improves the accuracy and efficiency of rumor detection by weighted summation, which makes the model more targeted to learn information from different sources.

4. Experiments

4.1. Datasets

We selected the BuzzFeed dataset [40] as well as manually crawled and constructed the PolitiFact dataset for our experiments and evaluations, which are recognized as the public benchmark datasets for fake news detection. The BuzzFeed dataset contains the news published by nine news organizations on the Facebook platform about the 2016 U.S. election and their truthfulness labels, where the number of fake-news articles is 355, and the number of true news articles is 1247. The PolitiFact dataset, sourced from FakeNewsNet, comprises news articles collected from the fact-checking website PolitiFact, labeled for authenticity by professionals. It includes news content, social context, and dynamic information. To facilitate experimentation and comparison, samples labeled as "mixed true and false" were treated as fake news, where the number of fake-news articles is 112, and the number of real-news articles is 463. This experiment only uses the news text information and corresponding labels of these two datasets. Specific statistics for all datasets are shown in the Table 1.

Table 1. The specific statistics of the dataset.

Dataset	Fake	Real	Total
BuzzFeed	355	1247	1604
PolitiFact	112	463	575

4.2. Setup

The training and validation sets and the test set are divided in the ratio of 7:1:2. Hardware resources used are Intel(R) Xeon(R) Gold 6138 CPU @ 2.00 GHz, GPU 4090Ti, made in Super Micro Computer, Inc. (Beijing, China). The sentence-level vector output of the last layer of the DeBERTa model is used as a representation of the textual features with a dimension of 768. The hidden dimension of the intermediate fully connected layer is 64. Each fully connected layer is followed by a dropout layer with a drop rate of 0.5. The batch size is set to 8, the optimizer uses Adam, and the initial learning rate is set to 1×10^{-5}, the

value of weight decay is set to 1×10^{-4}. The prefix prompt length of P-Tuning V2 is set to 128, and the learning rate of RumorLLM is set to 2×10^{-2}. The train epoch is 50.

The statistics of the datasets after RumorLLM data augmentation are shown in Table 2. BuzzFeed had a training set of 1122, augmented with RumorLLM data to add 355 fake news stories, with a final training set of 1477, a validation set of 161 , and a test set of 321. PolitiFact had a training set of 402, augmented with RumorLLM data to add 224 fake-news stories, with a final training set of 626, the number of validation sets is 58, and the number of test sets is 115. For the fairness of comparison, the same test set is used for all baselines and the proposed model, and no data augmentation was used on the test set.

Table 2. The statistics of the data set after RumorLLM data augmentation.

Dataset	TrainSet	ValidSet	TestSet	Augmentation	Augmentation TrainSet
BuzzFeed	1122	161	321	355	1477
PolitiFact	402	58	115	224	626

4.3. Evaluation Metrics

In this paper, we use accuracy, precision, recall, and F1 score to measure the performance of the model. Accuracy indicates the percentage of correct predictions. Precision indicates the percentage of predictions that are correct when the model predicts a positive sample. Recall is the result of the percentage of all positive samples that the model predicts correctly. The F1 score is the weighted summed average of precision and recall. The formula for evaluating the metrics is shown in Equations (7)–(10).

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

where T and F denote the correctness of the prediction. They indicate the correct and incorrect predictions. P and N denote the prediction categories of real news and fake news, respectively, and the result of summing these four values is the total number of samples.

In addition to this, in binary classification, the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) is a key metric for assessing model performance. This metric focuses on the balance between True Positive Rate (TPR) and False Positive Rate (FPR) across different decision thresholds.

ROC Curve: The ROC curve is plotted with FPR on the x-axis and TPR on the y-axis. TPR, also known as recall, represents the proportion of correctly predicted positives among actual positives. FPR indicates the proportion of incorrectly predicted positives among actual negatives.

AUC (Area Under the Curve): The AUC is the area under the ROC curve, a value ranging from 0 to 1, quantifying the model's classification performance. A higher AUC suggests superior performance across various decision thresholds.

ROC-AUC is particularly valuable when dealing with imbalanced datasets, as it remains robust against uneven class distributions, providing a comprehensive evaluation of model performance.

Independence from Class Distribution: ROC-AUC computation is not contingent on the actual distribution of class labels. In imbalanced datasets, where one class significantly outnumbers the other, other evaluation metrics may be influenced, whereas ROC-AUC

reflects the model's ability to classify positive and negative instances independently of their distribution.

Balancing TPR and FPR: The graphical representation of the ROC curve illustrates the trade-off between TPR and FPR. In cases of class imbalance, where the model may be biased towards the dominant class, TPR and FPR dynamics are visually apparent. ROC-AUC synthesizes these considerations, offering a holistic performance measure.

In binary classification, TPR and FPR are defined as follows:

$$\text{TPR} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{11}$$

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}} \tag{12}$$

ROC-AUC is calculated either by integrating the area under the ROC curve or approximating it using the trapezoidal rule:

$$\text{ROC-AUC} \approx \frac{1}{2}(\text{TPR}_1 \times \text{FPR}_2 + \text{TPR}_2 \times \text{FPR}_3 + \ldots) \tag{13}$$

Here, TPR_i and FPR_i represent TPR and FPR for the i-th positive instance in the sorted predicted probabilities. This calculation considers different thresholds, making it suitable for datasets with varying class distributions.

4.4. Baselines

Text-RF [41]: By focusing on the language style, text complexity, and psychological aspects of the text, they analyzed the news text, extracted 120 kinds of features, and adopted Random Forest as the classifier.

LR-Bias [42]: Researchers extract the LIWC features, readability features, and source bias features of news texts and use the logistic regression model to detect fake news.

Ternion [43]: a novel solution for detecting the authenticity of news on social media using natural-language processing techniques. The proposed scheme consists of three steps: stance detection, author credibility verification, and machine learning-based classification.

EANN [44]: EANN uses TEXT-CNN to extract news text features and VGG19 to extract news image features. For the sake of fairness in the comparison, only the part of the EANN that deals with textual features is considered.

SpotFake [45]: SpotFake mainly consists of a text encoder, an image feature extractor, a model fusion layer, and an output layer. It is designed with the goal of determining the authenticity of the news by integrating text and images through deep-learning models. In the experiments of this paper, the image feature extractor is removed to ensure the consistency of the comparison for a fair comparison. The model performs the final judgment of truthfulness through the output layer.

4.5. Results and Analysis

As shown in Table 3, the proposed model presented by the authors for the BuzzFeed dataset shows remarkable improvements in various performance metrics. While its accuracy reaches 82.55%, outperforming TEXT-RF (73.83%), LR-Bias (78.82%), Ternion (74.77%), EANN (73.21%), and SpotFake (79.44%), the increase in accuracy is not particularly substantial. However, the use of RumorLLM for data augmentation resulted in a significant improvement in F1 score and AUC-ROC values. Precision is particularly high at 0.70, outperforming other models, including LR-Bias (0.66). Furthermore, the proposed model achieves a robust recall of 0.5833, indicating a comprehensive performance improvement compared to SpotFake's 0.3452. On the F1 score, the proposed model significantly outperforms TEXT-RF (0.2075) and SpotFake (0.4677), reaching a score of 0.6364. In particular, on the AUC-ROC metric, the proposed model achieves a remarkable value of 0.8675, significantly outperforming SpotFake (0.8568) and LR-Bias (0.8176).

Table 3. Performance of baseline models and proposed model on the BuzzFeed dataset.

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
TEXT-RF	0.7383	0.5000	0.1310	0.2075	0.6230
LR-Bias	0.7882	0.6600	0.3929	0.4925	0.8176
Ternion	0.7477	0.5146	0.6310	0.5668	0.8051
EANN	0.7321	0.4900	0.5833	0.5326	0.7831
SpotFake	0.7944	0.7250	0.3452	0.4677	0.8568
Proposed	0.8255	0.7000	0.5833	0.6364	0.8675

As can be seen in Table 4, the proposed model introduced by the authors for the PolitiFact dataset shows remarkable progress in various performance metrics. While its accuracy is 93.91%, outperforming TEXT-RF (78.26%), LR-Bias (85.22%), Ternion (86.09%), EANN (82.61%), and SpotFake (86.96%), the increase in accuracy is not particularly substantial. However, the integration of RumorLLM for data augmentation resulted in significant improvements in F1 score and AUC-ROC values. In particular, the precision of the proposed model is 0.8519, outperforming other models, including LR-Bias (0.7647). In addition, the proposed model achieves a robust recall of 0.8846, indicating a significant improvement in performance compared to SpotFake's recall of 0.4615. In terms of F1 score, the proposed model significantly outperforms TEXT-RF (0.1429) and SpotFake (0.6154), achieving a score of 0.8679. Of particular note is the AUC-ROC metric, where the proposed model achieves a remarkable value of 0.9233, significantly outperforming SpotFake (0.8844) and LR-Bias (0.8619).

Table 4. Performance of baseline models and proposed model on the PolitiFact dataset.

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
TEXT-RF	0.7826	1.000	0.0769	0.1429	0.6791
LR-Bias	0.8522	0.7647	0.5000	0.6047	0.8619
Ternion	0.8609	0.7500	0.5769	0.6522	0.8429
EANN	0.8261	0.6154	0.6154	0.6154	0.8879
SpotFake	0.8696	0.9231	0.4615	0.6154	0.8844
Proposed	0.9391	0.8519	0.8846	0.8679	0.9233

It is crucial to highlight that, despite a modest increase in accuracy, the proposed model's substantial enhancements in F1 score and AUC-ROC effectively address the challenges posed by the imbalanced dataset. The high AUC-ROC score underscores the model's exceptional ability to handle scenarios with disparate positive and negative sample proportions. This further solidifies the proposed model's superiority in addressing imbalanced datasets, as AUC-ROC provides a comprehensive evaluation that is less influenced by data distribution. The model's proficiency in capturing the relationship between positive and negative classes in imbalanced data emphasizes its excellence not only in overall performance but also in accurately assessing discriminative capacity, presenting a reliable solution for handling such challenges.

4.6. Ablation Study

As part of the ablation study, we investigated the impact of data augmentation, specifically the use of RumorLLM, on the performance of the proposed model for the BuzzFeed dataset. The results, presented in Figures 2 and 3, show the performance contrast between the model without data augmentation and the augmented data by the proposed model.

On the BuzzFeed dataset, without data augmentation, the model achieved an accuracy of 75.70%, a precision of 55.36%, a recall of 36.90%, an F1 score of 44.29%, and an AUC-ROC of 82.34%. Subsequently, with the integration of RumorLLM for data augmentation, the proposed model showed significant improvements in all metrics. The augmented proposed model showed an accuracy of 82.55%, a precision of 70.00%, a recall of 58.33%, an F1 score of 63.64%, and an AUC-ROC of 86.75%. Notably, these results underscore the

significant positive impact of using RumorLLM for data augmentation, contributing to significant improvements in precision, recall, F1 score, and AUC-ROC. The results highlight the effectiveness of data augmentation in refining the model's performance and emphasize its role in capturing nuanced patterns and relationships within the BuzzFeed dataset.

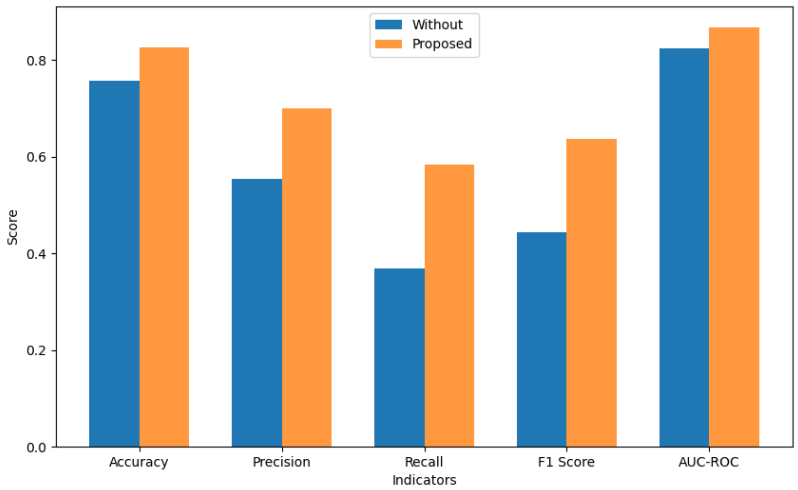

Figure 2. Performance comparison of the proposed model on the BuzzFeed dataset with and without data augmentation.

On the Politifact dataset, the model achieved the following scores without data augmentation: an accuracy of 91.30%, a precision of 86.36%, a recall of 73.08%, an F1 score of 79.17%, and an AUC-ROC of 90.86%. However, after incorporating RumorLLM for data augmentation, the proposed model demonstrated significant improvements across all metrics. The augmented proposed model achieved an accuracy of 93.91%, a precision of 85.19%, a recall of 88.46%, an F1 score of 86.79%, and an AUC-ROC of 92.33%. These results highlight the substantial positive impact of using RumorLLM for data augmentation, resulting in significant improvements in precision, recall, F1 score, and AUC-ROC. The findings underscore the effectiveness of data augmentation in refining the model's performance and emphasize its role in capturing nuanced patterns and relationships within the Politifact dataset.

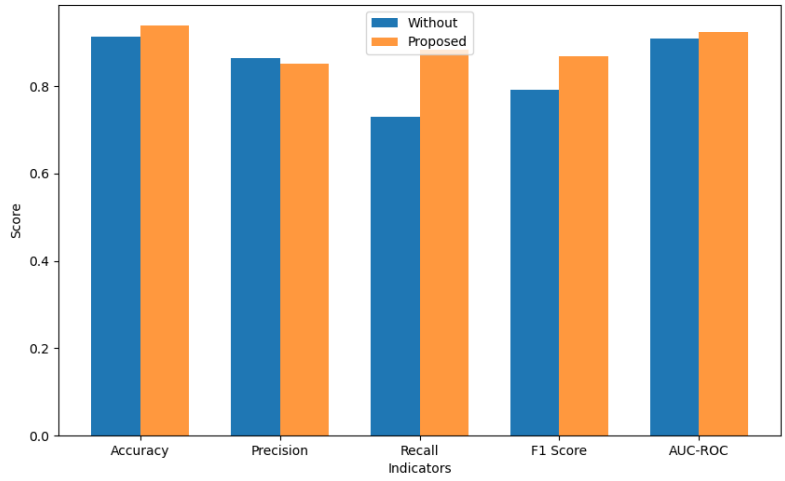

Figure 3. Performance comparison of the proposed model on the PolitiFact dataset with and without data augmentation.

4.7. Case Study

Case: *Former President George H.W. Bush said he will be voting for Hillary Clinton at a reception for the Points of Light Foundation in Kennebunkport, Maine, on Monday, a source told ABC News. Bush made the declaration in front of about 40 people who were in attendance at the reception, according to the source. Kathleen Hartington Kennedy Townsend, a former Maryland lieutenant governor and daughter of Robert F. Kennedy, first posted on Facebook Monday about the 41st president's apparent intention to vote for Clinton, sharing a photo of herself and Bush, and writing, " The President told me he's voting.*

The purpose of this case study is to evaluate the effectiveness of a rumor generated by RumorLLM. Through an artificial factual study of a deceptive rumor claiming that former President George H.W. Bush endorsed Hillary Clinton, we find that the lack of credible sources and official validation of the rumor is strong evidence that the rumor was fabricated. The fact that this fake news was generated by RumorLLM highlights the model's nuanced grasp of the context of political news, as well as its superior ability to mimic the unique writing style associated with political rumors. This case serves as a profound reminder that when reviewing language model output, especially in a domain as nuanced as politics, RumorLLM can also generate rumors with content as well as writing styles that can be virtually faked. It highlights the critical need for continuous model evaluation, enhanced interpretability features, and ethical considerations to curb the unintentional spread of misinformation when deploying such advanced AI systems.

5. Conclusions

In conclusion, this paper presents a novel approach to addressing the challenges posed by false information, rumors, and misleading content in the digital age. The proliferation of such content on the Internet and social media platforms has become a significant concern, and effective methods for detecting and combating fake news are essential. We propose the construction of a rumor-generating large language model called "Rumor Large Language Models" (RumorLLM). This model is created by finetuning large language models with rumor writing styles and content. By leveraging RumorLLM and prompt engineering, the authors demonstrate a method for data enhancement in small categories, which helps to address the issue of category imbalance in real-world fake-news datasets.

This paper implements RumorLLM, which fills the gap in large-scale language models dedicated to fake news detection. The main advantage of this approach is the introduction of RumorLLM, which is specifically tailored to generate rumors. Leveraging large language models' semantic understanding and generation capabilities, RumorLLM captures the writing style and content characteristics of rumors, leading to more accurate and contextually relevant rumor generation. We also employ prompt engineering and data enhancement techniques to address category imbalance in real-world fake-news datasets. By generating diversified samples for the minority class, the approach improves the accuracy and efficiency of fake news detection, providing a more comprehensive evaluation of model performance.

By assembling RumorLLM with state-of-the-art classification models and evaluating the results on real datasets, the authors demonstrate the effectiveness of their approach. The experimental results on the BuzzFeed dataset show that the proposed model outperforms baseline models in various evaluation metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. The model's ability to handle imbalanced datasets is particularly noteworthy, as it significantly improves the F1 score and AUC-ROC, providing a comprehensive evaluation of its performance. The paper also identifies potential avenues for future research. Additional enhancements to RumorLLM can be explored, including investigating the model's interpretability. Further evaluation of the model's performance on diverse datasets is also recommended. Additionally, the proposed approach can be extended to address new challenges that may arise in the ever-evolving landscape of false information and misinformation.

5.1. Limitations

However, there are some limitations and future research directions to consider. For example, the current RumorLLM is limited to the generation of rumors for the plain text category, which is relatively weak for the multimodal (including audio, image, video, etc.) types of news that are currently proliferating on the Internet. Another limitation is the potential ethical implications of using RumorLLM. As a tool for generating rumors, there is a risk that it could be misused to spread false information or contribute to disinformation campaigns. It is important to ensure the responsible use of RumorLLM and consider ethical guidelines and safeguards to prevent malicious or harmful applications.

5.2. Research Potentials

Using RumorLLM for data enhancement offers more options than traditional methods (e.g., randomly adding, deleting, or changing the word order of the original text). Traditional methods always leave the original data unchanged no matter how they are changed, but using RumorLLM for data enhancement allows us to obtain more diverse samples by constantly finetuning the prompts (e.g., reframing, associating, retelling, etc.). RumorLLM is specifically designed for rumor generation and can generate text that better fits the characteristics of the rumor, improving the model's ability to generalize to real data. By generating rumor text, RumorLLM provides richer and more diversified data enhancement methods to capture complex linguistic structures and meanings, and the generated data helps the model to learn and distinguish rumors from real information more accurately and improves the performance of rumor detection. As an innovative data enhancement method, RumorLLM provides a platform for rumor-detection research and promotes the exploration of more problems and the proposal of new methods on rumor generation and detection.

At the same time, RumorLLM possesses vast research potential and can contribute significantly to the fields of fake news detection, explanatory research, tackling emerging challenges in combating misinformation, and other related domains. Through further research and application, we can continuously enhance the capabilities of fake news detection and combat misinformation effectively, therefore upholding the integrity and fairness of information dissemination. The research potential of RumorLLM is extensive and can be elaborated as follows:

1. Enhanced accuracy in fake news detection: RumorLLM's specialized design for generating rumors can improve the accuracy of fake-news-detection algorithms by simulating and generating rumors more accurately.
2. Interpreting large language models: Studying RumorLLM's decision-making process can provide insights into how large language models generate deceptive content, enhancing their interpretability and transparency.
3. Addressing emerging challenges: RumorLLM can be applied to tackle new forms of misinformation, such as deepfakes and coordinated disinformation campaigns, contributing to ongoing research in combating false information.
4. Generalization to other domains: The methods and techniques of RumorLLM can be extended to areas like natural-language processing, sentiment analysis, and social media mining, improving the detection of deceptive content in diverse contexts.

5.3. Future Work

The main areas of future work in response to the research in this paper are as follows: First, the lack of support for multimodal content poses a challenge in accurately detecting and debunking misinformation across different media types. Future research should aim to extend RumorLLM's capabilities to encompass multimodal content generation and detection. This would require incorporating techniques for analyzing and synthesizing audio, image, and video-based rumors, as well as developing novel methods for detecting and debunking multimodal misinformation. Exploring the interpretability of RumorLLM is important to understand its decision-making process and enhance transparency. Evaluating

the method on a variety of datasets from different sources and domains allows for a comprehensive assessment of its generalizability and exploration of how to quantitatively assess the effectiveness of RumorLLM-generated rumor generation compared to original rumor generation. In addition, how can we consider enhancing RumorLLM's ability to interpret fake-news features (e.g., quantitative assessment features, features such as news text sentiment, positional tendencies, etc.) so that RumorLLM can more accurately generate rumor-style fake news? Additionally, future research should focus on addressing emerging challenges like deepfake content, evolving writing styles, and coordinated disinformation campaigns to further combat fake news.

In summary, this paper contributes to the field of fake news detection by proposing a novel approach using rumor-generating large language models. The experimental results demonstrate the effectiveness of the proposed model in handling the challenges posed by false information, and future work can build upon these findings to further advance the field.

Author Contributions: Conceptualization, J.L. and X.S.; methodology, J.L.; validation, W.L. and L.Z.; formal analysis, J.L. and L.L.; investigation, J.L., W.L. and X.Y.; resources, X.S. and Y.W.; data curation, Y.W. and X.Y.; writing—original draft preparation, J.L.; writing—review and editing, J.L., X.S. and X.Y.; project administration, J.L. and X.S.; funding acquisition, X.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (No.2021YFB3101100). This work was sponsored by the National Natural Science Foundation of China, General Program with grant number (No.62272352). This research is supported in part by the Humanities and Social Sciences of Ministry of Education Planning Fund (No.21YJAZH073).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to copyright.

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

References

1. Capuano, N.; Fenza, G.; Loia, V.; Nota, F.D. Content-Based Fake News Detection with Machine and Deep Learning: A Systematic Review. *Neurocomputing* **2023**, *530*, 91–103. [CrossRef]
2. Zhou, K.; Shu, C.; Li, B.; Lau, J.H. Early rumour detection. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 2–7 June 2019; Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 1614–1623. Available online: https://aclanthology.org/N19-1163 (accessed on 3 February 2024).
3. Liu, Y.; Xu, S. Detecting rumors through modeling information propagation networks in a social media environment. *IEEE Trans. Comput. Soc. Syst.* **2016**, *3*, 46–62. [CrossRef]
4. Sampson, J.; Morstatter, F.; Wu, L.; Liu, H. Leveraging the implicit structure within social media for emergent rumor detection. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, ser. CIKM '16, Indianapolis, IN, USA, 24–28 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 2377–2382. . [CrossRef]
5. 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] [PubMed]
6. Varshini, U.S.S.; Sree, R.P.; Srinivas, M.; Subramanyam, R.B.V. Rdgt-gan: Robust distribution generalization of transformers for covid-19 fake news detection. *IEEE Trans. Comput. Soc. Syst.* **2023**, *11*, 1–15. [CrossRef]
7. Hu, Y.; Ju, X.; Ye, Z.; Khan, S.; Yuan, C.; Lai, Q.; Liu, J. Early rumor detection based on data augmentation and pre-training transformer. In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Virtual, 26–29 January 2022; pp. 152–158.
8. Zhou, X.; Jain, A.; Phoha, V.V.; Zafarani, R. Fake news early detection: A theory-driven model. *Digit. Threat.* **2020**, *1*, 1–25. . [CrossRef]
9. Amjad, M.; Sidorov, G.; Zhila, A. Data augmentation using machine translation for fake news detection in the Urdu language. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Eds.; European Language Resources Association: Marseille, France, 2020; pp. 2537–2542. Available online: https://aclanthology.org/2020.lrec-1.309 (accessed on 3 February 2024).

10. Wei, J.; Zou, K. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; Inui, K., Jiang, J., Ng, V., Wan, X., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 6382–6388. Available online: https://aclanthology.org/D19-1670 (accessed on 3 February 2024).
11. Bhattacharjee, S.; Maity, S.; Chatterjee, S. Addressing class imbalance in fake news detection with latent space resampling. In *Computational Intelligence in Pattern Recognition*; Das, A.K., Nayak, J., Naik, B., Vimal, S., Pelusi, D., Eds.; Springer Nature: Singapore, 2023; pp. 427–438.
12. Prasetijo, A.B.; Isnanto, R.R.; Eridani, D.; Soetrisno, Y.A.A.; Arfan, M.; Sofwan, A. Hoax detection system on indonesian news sites based on text classification using svm and sgd. In Proceedings of the 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 18–19 October 2017; pp. 45–49.
13. Granik, M.; Mesyura, V. Fake news detection using naive bayes classifier. In Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv, Ukraine, 29 May–2 June 2017; pp. 900–903.
14. Lyu, S.; Lo, D.C.-T. Fake news detection by decision tree. In Proceedings of the 2020 SoutheastCon, Raleigh, NC, USA, 28–29 March 2020; pp. 1–2.
15. Castillo, C.; Mendoza, M.; Poblete, B. Information credibility on twitter. In Proceedings of the 20th International Conference on World Wide Web, ser. WWW '11, Hyderabad, India, 28 March–1 April 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 675–684. . [CrossRef]
16. Ruchansky, N.; Seo, S.; Liu, Y. Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, ser. CIKM '17, Singapore, 6–10 November 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 797–806. . [CrossRef]
17. Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.-F.; Cha, M. Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, ser. IJCAI'16, New York, NY, USA, 9–15 July 2016; AAAI Press: Washington, DC, USA, 2016; pp. 3818–3824.
18. Tan, L.; Wang, G.; Jia, F.; Lian, X. Research status of deep learning methods for rumor detection. *Multimed. Tools Appl.* **2022**, *82*, 2941–2982. . [CrossRef] [PubMed]
19. Zhang, P.; Ran, H.; Jia, C.; Li, X.; Han, X. A lightweight propagation path aggregating network with neural topic model for rumor detection. *Neurocomputing* **2021**, *458*, 468–477. [CrossRef]
20. Yu, F.; Liu, Q.; Wu, S.; Wang, L.; Tan, T. A convolutional approach for misinformation identification. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, ser. IJCAI'17, Melbourne, Australia, 19–25 August 2017; AAAI Press: Washington, DC, USA, 2017; pp. 3901–3907.
21. Vaibhav, V.; Mandyam, R.; Hovy, E. Do sentence interactions matter? Leveraging sentence level representations for fake news classification. In Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13), Hong Kong, China, 4 November 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 134–139. Available online: https://aclanthology.org/D19-5316 (accessed on 3 February 2024).
22. Alzanin, S.M.; Azmi, A.M. Rumor detection in arabic tweets using semi-supervised and unsupervised expectation–maximization. *Knowl.-Based Syst.* **2019**, *185*, 104945. [CrossRef]
23. Ma, J.; Gao, W.; Wong, K.-F. Detect rumors on twitter by promoting information campaigns with generative adversarial learning. In Proceedings of the World Wide Web Conference, ser. WWW '19, San Francisco, CA, USA, 13–17 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 3049–3055. . [CrossRef]
24. Su, T.; Macdonald, C.; Ounis, I. Ensembles of recurrent networks for classifying the relationship of fake news titles. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR'19, Paris, France, 21–25 July 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 893–896. . [CrossRef]
25. Zhou, H.; Ma, T.; Rong, H.; Qian, Y.; Tian, Y.; Al-Nabhan, N. Mdmn: Multi-task and domain adaptation based multi-modal network for early rumor detection. *Expert Syst. Appl.* **2022**, *195*, 116517. . [CrossRef]
26. Cao, J.; Qi, P.; Sheng, Q.; Yang, T.; Guo, J.; Li, J.; *Exploring the Role of Visual Content in Fake News Detection*; Springer International Publishing: Cham, Switzerland, 2020; pp. 141–161. . [CrossRef]
27. Qi, P.; Cao, J.; Yang, T.; Guo, J.; Li, J. Exploiting multi-domain visual information for fake news detection. In Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019; pp. 518–527.
28. Wu, Y.; Zhan, P.; Zhang, Y.; Wang, L.; Xu, Z. Multimodal fusion with co-attention networks for fake news detection. In Proceedings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Online, 1–6 August 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 2560–2569. Available online: https://aclanthology.org/2021.findings-acl.226 (accessed on 3 February 2024).
29. Jin, Z.; Cao, J.; Guo, H.; Zhang, Y.; Luo, J. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In Proceedings of the 25th ACM International Conference on Multimedia, ser. MM '17, Mountain View, CA, USA, 23–27 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 795–816. . [CrossRef]
30. Ran, H.; Jia, C. Unsupervised cross-domain rumor detection with contrastive learning and cross-attention. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023. Available online: https://api.semanticscholar.org/CorpusID:257636865 (accessed on 3 February 2024).

31. Song, C.; Ning, N.; Zhang, Y.; Wu, B. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. *Inf. Process. Manag.* **2021**, *58*, 102437. [CrossRef]
32. Qian, S.; Wang, J.; Hu, J.; Fang, Q.; Xu, C. Hierarchical multi-modal contextual attention network for fake news detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR '21, Virtual, 11–15 July 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 153–162. . [CrossRef]
33. Hossain, M.M.; Awosaf, Z.; Prottoy, M.S.H.; Alvy, A.S.M.; Morol, M.K. Approaches for improving the performance of fake news detection in bangla: Imbalance handling and model stacking. In Proceedings of the International Conference on Fourth Industrial Revolution and Beyond 2021, Dhaka, Bangladesh, 10–11 December 2021; Hossain, S., Hossain, M.S., Kaiser, M.S., Majumder, S.P., Ray, K., Eds.; Springer Nature: Singapore, 2022; pp. 723–734.
34. Salah, I.; Jouini, K.; Korbaa, O. Augmentation-based ensemble learning for stance and fake news detection. In *Advances in Computational Collective Intelligence*; Bădicxax, C., Treur, J., Benslimane, D., Hnatkowska, B., Krótkiewicz, M., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 29–41.
35. Naveed, H.; Khan, A.U.; Qiu, S.; Saqib, M.; Anwar, S.; Usman, M.; Barnes, N.; Mian, A.S. A comprehensive overview of large language models. *arXiv* **2023**, arXiv:2307.06435. Available online: https://api.semanticscholar.org/CorpusID:259847443 (accessed on 3 February 2024).
36. Beguš, G.; Dąbkowski, M.; Rhodes, R. Large linguistic models: Analyzing theoretical linguistic abilities of llms. *arXiv* **2023**, arXiv:2305.00948.
37. Hu, J.E.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Chen, W. Lora: Low-rank adaptation of large language models. *arXiv* **2021**, arXiv:2106.09685. Available online: https://api.semanticscholar.org/CorpusID:235458009 (accessed on 3 February 2024).
38. Liu, X.; Ji, K.; Fu, Y.; Tam, W.; Du, Z.; Yang, Z.; Tang, J. P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Dublin, Ireland, 22–27 May 2022; Muresan, S., Nakov, P., Villavicencio, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; pp. 61–68. Available online: https://aclanthology.org/2022.acl-short.8 (accessed on 3 February 2024).
39. He, P.; Liu, X.; Gao, J.; Chen, W. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv* **2020**, arXiv:2006.03654. Available online: https://api.semanticscholar.org/CorpusID:219531210 (accessed on 3 February 2024).
40. Potthast, M.; Kiesel, J.; Reinartz, K.; Bevendorff, J.; Stein, B. A stylometric inquiry into hyperpartisan and fake news. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; Gurevych, I., Miyao, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 231–240. Available online: https://aclanthology.org/P18-1022 (accessed on 3 February 2024).
41. Shrestha, A.; Spezzano, F. Textual characteristics of news title and body to detect fake news: A reproducibility study. In Proceedings of the Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, 28 March–1 April 2021; Proceedings, Part II; Springer: Berlin/Heidelberg, Germany, 2021; pp. 120–133. . [CrossRef]
42. Shrestha, A.; Spezzano, F.; Gurunathan, I. Multi-modal analysis of misleading political news. In Proceedings of the Disinformation in Open Online Media: Second Multidisciplinary International Symposium, MISDOOM 2020, Leiden, The Netherlands, 26–27 October 2020; Proceedings; Springer: Berlin/Heidelberg, Germany, 2020; pp. 261–276. . [CrossRef]
43. Islam, N.; Shaikh, A.; Qaiser, A.; Asiri, Y.; Almakdi, S.; Sulaiman, A.; Moazzam, V.; Babar, S.A. Ternion: An autonomous model for fake news detection. *Appl. Sci.* **2021**, *11*, 9292. [CrossRef]
44. Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; Gao, J.; "Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery, London, UK, 19–23 August 2018; Data Mining, ser. KDD '18; Association for Computing Machinery: New York, NY, USA, 2018; pp. 849–857. . [CrossRef]
45. Singhal, S.; Shah, R.R.; Chakraborty, T.; Kumaraguru, P.; Satoh, S. Spotfake: A multi-modal framework for fake news detection. In Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), Singapore, 11–13 September 2019; pp. 39–47.