Enhancing multilingual fake news detection through LLM-Based data augmentation

Razieh Chalehchaleh, Reza Farahbakhsh, Noel Crespi

Enhancing Multilingual Fake News Detection through LLM-Based Data Augmentation

Razieh Chalehchaleh, Reza Farahbakhsh, and Noel Crespi

Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France
{razieh.chalehchaleh, reza.farahbakhsh, noel.crespi}@telecom-sudparis.eu

Abstract. The rapid growth of online news consumption has intensified the spread of misinformation, underscoring the critical need for effective fake news detection methods. Despite significant advancements in this area, the scarcity and inadequacy of high-quality labeled datasets necessary for training effective detection models remains a major challenge. In this paper, we introduce a novel approach to address this issue by leveraging large language models (LLMs) for data augmentation. Specifically, we employ Llama 3 to generate multiple synthetic news samples per original article, enriching existing fake news datasets to enhance fake news detection. We explore various augmentation strategies like different augmentation rates, random or similarity-based subsampling, and selectively augmenting data from specific classes to optimize the augmented datasets to train better classifiers. We evaluate the efficacy of our approach using BERT-based classifiers on two multilingual datasets. Our findings reveal notable improvements particularly when augmenting only the fake class with rate 1.

Keywords: Multilingual Fake News Detection, Large Language Models (LLMs), Data Augmentation

1 Introduction

In recent years, rapid technological advancements and their widespread adoption have reshaped the landscape of news consumption. This transformation has triggered a notable surge in the usage of online platforms for accessing news content. A 2023 survey[1] found that 86% of U.S. adults frequently or occasionally access news via digital devices with 50% at least sometimes obtaining news from social media. Another 2023 study [2] highlights that 40% of Americans who use social media to access news are concerned about inaccuracies—encompassing issues such as unverified facts, misinformation, the spread of fake news, and reliance on unreliable sources. The COVID-19 pandemic accelerated the spread of fake news, emphasizing the need to combat false information. A poll involving 806

[1] https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet
[2] https://www.pewresearch.org/short-reads/2024/02/07/many-americans-find-value-in-getting-news-on-social-media-but-concerns-about-inaccuracy-have-risen

U.S. physicians[3] reveals that misinformation is perceived as a significant issue with 44% of the respondents indicating that more than half of the COVID-19 information they encountered from patients was medically inaccurate. Moreover, 72% reported that misinformation aggravated difficulties in treating patients.

While there is no universally accepted definition of "fake news" in the literature, Allcott and Gentzkow[3] offer a widely recognized definition, describing it as "news articles that are intentionally and verifiably false and could mislead readers." Extensive research has been devoted to addressing the critical task of fake news detection. A thorough review of research and methods for detecting and preventing the spread of fake news on online social networks is presented in [2]. One of the most critical aspects of training supervised models and enhancing their efficiency is the training data. The lack of labeled data, particularly in under-resourced languages, poses a significant challenge. Although several datasets are available for fake news detection, they often face limitations such as size, modality, granularity, and becoming outdated over time [13]. To tackle the issue of limited data availability, researchers have utilized diverse techniques to augment existing fake news datasets. These methods aim to enhance both the volume and diversity of training data, thereby improving model performance without the need for further data collection efforts. However, the outcomes of these augmentation approaches in fake news detection have been varied. While some studies have shown improvements in overall outcomes [14, 17], others have reported insignificant, negative, or mixed effects on the results [4, 16, 5, 15, 1].

Large Language Models (LLMs) have revolutionized the field of natural language processing (NLP) [27, 20]. These models demonstrate robust abilities to comprehend natural language and tackle complex tasks, particularly through text generation [26]. Leveraging LLMs for data augmentation presents a valuable use case as they can sometimes produce high-quality synthetic data that exceeds human-curated ones. This approach helps address the scarcity of human-annotated data and permits the expansion of training datasets without proportionally increasing computational costs. Additionally, the use of synthetic data can substantially reduce data collection expenses and energy consumption [11].

In this study, we aim to address the challenges posed by the scarcity and inadequacy of labeled fake news datasets by augmenting the existing fake news datasets with synthetic news generated by LLMs. We utilize Llama 3[4], an open-source LLM developed by META GenAI, for fake news data augmentation. We will instruct the model to generate multiple news samples per original article, maintaining the same labels as the original pieces. To optimize our dataset augmentation for fake news detection, we will explore various strategies. After generating the augmented datasets, we will train BERT (Bidirectional Encoder Representations from Transformers) [10] classifiers for the news classification. Our strategies will include experimenting with different augmentation rates—the number of additional news samples generated per original sample, using random

[3] https://debeaumont.org/news/2023/physician-poll-medical-misinformation-is-harming-patients

[4] https://llama.meta.com/llama3

or similarity-based subsampling, and selectively augmenting data from specific classes (both fake and real, only fake, or only real). A schematic overview of our approach is provided in Fig. 1.

Our results on the two multilingual real-world datasets TALLIP [9] and MM-COVID [18] underscored the effectiveness of data augmentation in enhancing model performance. Notably, when augmenting only the fake class with one synthetic sample per original article (rate 1 augmentation), the approach consistently outperformed others. For instance, in the TALLIP dataset, augmenting only the fake class led to a notable improvement in F1 scores, with English increasing by 7.7% and Hindi by 4.4%. In the MM-COVID dataset, augmenting only the fake class improved F1 scores by 1.5% for Spanish and 1.1% for both Hindi and English. While similarity-based sampling generally yielded higher F1 scores, the difference was less pronounced when augmenting only real or fake news, suggesting consistent quality across the generated samples.

The rest of the paper is organized as follows: Section 2 reviews related work. Section 3 details our augmentation techniques and classification method. Section 4 covers the datasets, evaluation setting, and results. Section 5 addresses limitations and future research directions. Section 6 summarizes our approach and findings.

2 Related Work

Extensive research has been dedicated to the crucial task of fake news detection. A comprehensive review of fake news research and the examination of current methods for detecting and preventing the spread of fake news on online social networks is presented in [2]. A critical factor in training supervised models and enhancing their efficiency is access to high-quality datasets. Although numerous datasets are available for fake news detection[12], they often suffer from limitations such as size, modality, granularity, and becoming outdated over time [13]. These challenges are particularly pronounced for under-resourced languages.

In the literature, numerous researchers have attempted to augment fake news datasets using various techniques, addressing the challenges of not having a sufficient amount of data. Data augmentation refers to utilizing innovative techniques to enhance model performance by expanding the volume and diversity of training data, without the need for additional data collection. Bayer et al. conducted a comprehensive survey of various augmentation methods in [6]. The results of these augmentation approaches in the field of fake news detection have shown inconsistency. While some studies have reported improvements in overall outcomes, many others have found insignificant, negative, or mixed effects on the results. Amjad et al. [4] employed an augmentation approach to train a fake news classifier in Urdu, using both manually annotated and machine-translated English datasets, but found no significant improvement in performance.

Pre-trained language models (PLMs) based on transformer architectures [25] have revolutionized the field of natural language processing (NLP). Trained on large-scale corpora, these models exhibit robust capabilities across a spectrum

of tasks. Models like BERT (Bidirectional Encoder Representations from Transformers) [10] and RoBERTa (Robustly optimized BERT approach) [19], developed using transformer architecture and self-attention mechanisms, have shown exceptional abilities in capturing linguistic patterns and semantic relationships. The introduction of these comprehensive context-aware representations has led to widespread adoption by researchers across various tasks and applications. Hua et al. [14] introduced a multimodal approach that used a BERT-based back-translation method to augment the dataset size. Their results demonstrate promising improvements. Keya et al. [17] introduced AugFake-BERT, which leverages BERT to generate synthetic texts, thereby expanding the dataset with augmented fake data. Their study revealed an improvement in classification outcomes when applying text augmentation to the minority class. Kapusta et al. [16] used various text data augmentation techniques, including synonym replacement, back translation, and reduction of function words, to optimize a text corpus for fake news classification, revealing significant differences in classifier outcomes between augmented and original corpora. Ashraf et al. [5] applied random word insertion or replacement using Word2Vec similarity search for news claim classification but found no improvement. Junior et al. [15] augmented an English dataset by replacing nouns with synonyms based on part of speech tagging and similarity thresholds, resulting in a decrease in validation and test accuracy when translated to Portuguese for BERT classification.

Large language models (LLMs) have gained significant attention for their impressive performance across various tasks ranging from general NLP to domain-specific applications, especially following the release of OpenAI's ChatGPT[5]. These models, known for their proficiency in language understanding and generation, are commonly recognized as transformer-based language models with hundreds of billions of parameters, trained on extensive textual corpora [22]. Minaee et al. provided an extensive review of LLMs in their work [20]. Ding et al. [11] conducted a comprehensive study on the potential of using LLMs for data augmentation in NLP. Ahlback et al. [1] investigated various data augmentation methods, including pre-trained LLMs like Vicuna [8] and ChatGPT, along with a word substitution technique. They found modest improvements in fake news classification performance but no statistical significance. In our earlier research [7], we explored various multilingual models and training scenarios for detecting multilingual fake news. Building on this foundation, the current research aims to address the challenge of insufficiently sized annotated fake news datasets by augmenting them with synthetic news generated by LLMs. We introduce a novel approach using Llama 3 for data augmentation, with the goal of training better-performing detection models.

3 Methodology

To address the challenge of a limited supply of labeled fake news datasets for effective training of detection models, our research endeavors to augment mul-

[5] https://openai.com/index/chatgpt/

tilingual datasets with additional synthetic news generated by large language models (LLMs). A schematic overview of our approach is provided in Fig. 1. First, all news articles will be translated into English to ensure a common environment. We will use a specific prompt on Llama 3 to generate N new samples per news item, retaining the original labels. To determine the optimal dataset augmentation settings various augmentation strategies, including different rates (using random or similarity-based subsampling) and selective class augmentation, will be tested. After augmentation, we will train BERT-based classifiers. Further details are elaborated in the subsequent subsections.

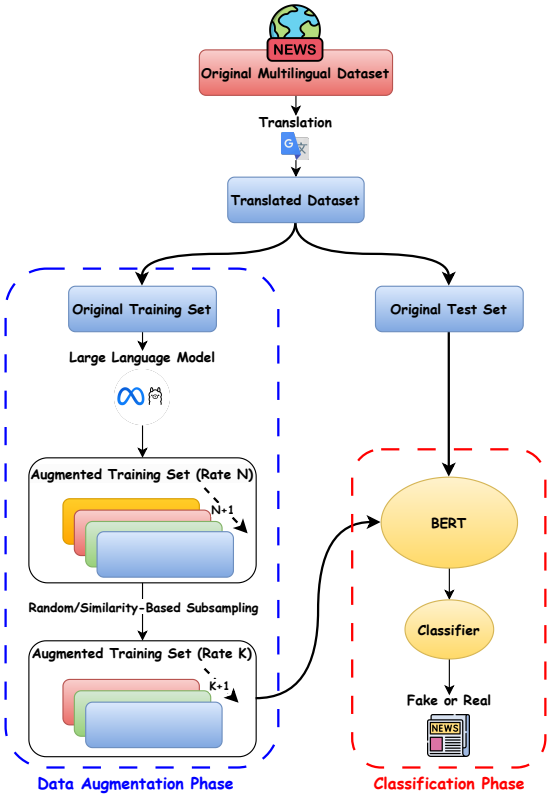

Fig. 1. A schematic overview of the proposed approach for enhancing fake news detection using LLM-based data augmentation

3.1 Data Augmentation with Llama 3

Large Language Models (LLMs) have demonstrated significant potential, excelling in complex reasoning tasks that require expert knowledge across various fields. Llama 3, developed by META GenAI, is a family of open-source pretrained and fine-tuned LLMs ranging from 8 billion to 70 billion parameters,

outperforming many existing models. For our task, we use the 8B instruction-tuned variant optimized for dialogue and chat use cases. Llama 3 builds on the strengths of its predecessors, Llama 1 [23] and Llama 2 [24]. To augment datasets with Llama 3, we use a carefully crafted prompt to generate paraphrased versions of each news item, retaining the original meaning and label. We employed prompt engineering and tested various structures to finalize the prompt used for generation. An example of the news generation is illustrated in Fig. 2. Further details on the selected prompt are provided in Section 4.2. We will explore different augmentation strategies to find the best settings for our fake news detection task, including varying augmentation rates (using random or similarity-based subsampling) and selectively augmenting specific classes (only fake, only real, or both).

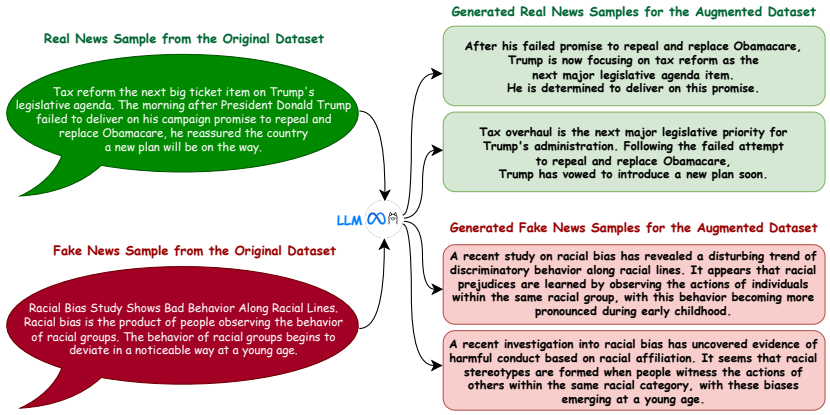

Fig. 2. An example of LLM-based synthetic news generation for dataset augmentation

Different Augmentation Rates The augmentation rate refers to the number of generated news samples per original news item. We aim to identify an optimal rate for the fake news dataset to achieve the best outcome without risking overfitting or introducing excessive noise. For this purpose, first, we generate N samples for each news item and then select K ($K \leq N$) subsamples, either randomly or based on similarity to form the K rated augmented dataset. The similarity-based approach uses Sentence-BERT (SBERT) [21] to obtain embeddings for generated samples and calculates their cosine similarity to the original news item's embedding to identify the most similar samples.

Different Classes to Be Augmented Another crucial factor that could impact the results is the choice of classes for augmentation. Specifically, given that real-world datasets are often imbalanced, prioritizing the augmentation of certain classes, such as fake news, may be advantageous. We will evaluate the effects

of three different augmentation strategies: augmenting both classes, augmenting only the fake class, and augmenting only the real class, to understand how these class-specific approaches impact detection accuracy.

3.2 BERT-Based Classification

After acquiring the augmented datasets, we utilize BERT (Bidirectional Encoder Representations from Transformers) [10] to generate text embeddings. BERT, pre-trained on large corpora with masked language modeling and next-sentence prediction, produces contextual embeddings through multiple transformer layers. A linear classification layer then maps these embeddings to binary labels (fake or real). The model is trained to minimize classification loss, enhancing its ability to accurately classify news as fake or real.

4 Experiments

In this section, we detail the datasets, experimental settings, and the results of our study. We first introduce the two multilingual datasets and experimental setup, then explore the key findings and the impact of different strategies.

4.1 Dataset

In our study, we leverage two multilingual datasets: TALLIP[6] [9] and MM-COVID[7] [18]. The TALLIP dataset encompasses five languages: English (en), Hindi (hi), Indonesian (id), Swahili (sw), and Vietnamese (vi). It includes articles from various domains, including business, education, technology, entertainment, politics, sports, and celebrity news. This dataset maintains a balance between the fake and real classes, with 490 fake and 490 real news articles for each language, yielding a total of 2,450 articles for each class. MM-COVID comprises multilingual COVID-19-related news; we focus on 6 languages, English (en), Spanish (es), French (fr), Hindi (hi), Italian (it), and Portuguese (pt). The statistics of the MM-COVID dataset used in our experiments can be seen in Table 1. English and Hindi are the two languages shared by both datasets.

4.2 Experimental Setting

We conducted our experiments using the computing resources available through Google Colaboratory. Our models are developed using PyTorch and the Hugging Face Transformers libraries. For translation, we employed Google Translate. For Llama 3, we utilize a directly quantized 4bit 8B instruction tuned model [8] optimized for dialogue use cases. We configure the max_new_tokens parameter to

[6] https://github.com/Arko98/TALLIP-FakeNews-Dataset
[7] https://github.com/bigheiniu/MM-COVID
[8] https://huggingface.co/unsloth/llama-3-8b-Instruct-bnb-4bit

Table 1. MM-COVID dataset statistics

Dataset	Language	Label	
		Fake	Real
MM-COVID	English (en)	2046	5081
	Spanish (es)	596	2271
	French (fr)	164	691
	Hindi (hi)	355	1205
	Italian (it)	107	938
	Portuguese (pt)	313	638
	Total	3581	10824

2000, with the temperature set to 0.6. For classification tasks, we employ the base uncased version of the BERT model. For text embedding generation before calculating the cosine similarity, we employ a SentenceTransformers [21] model[9] that maps sentences and paragraphs to a dense 768-dimensional vector space. We conducted training for our models over 10 epochs. The learning rate was configured to 2×10^{-5}, and the batch size was set to 28. During testing, 40% of each dataset was randomly selected, while the remaining 60% was allocated for training purposes. To augment the datasets using Llama 3, we employed a specific prompt to generate five paraphrased versions of news texts while preserving their original meanings and labels. The prompt is provided in the following box. In practice, each news text was substituted for `statement` in the template. An example of the paraphrased news generated using this prompt is shown in Fig. 2.

> Paraphrase the following text delimited by triple backquotes in 5 different ways. The generated texts must retain the exact same meaning as the original text. Enclose each paraphrased text within the following keyword pair: [BEGINPARA] ... [ENDPARA]. Only return the texts and no extra explanation.
>
> ‘‘‘`{statement}`‘‘‘
>
> NUMBERED GENERATED TEXTS:

4.3 Results

In this section, we will analyze the results of our experiments. The F1 results based on different augmentation classes and rates (with random subsampling) for the TALLIP and MM-COVID datasets using BERT-based classifiers are presented in Table 2. The top two F1 scores for each language are bolded in with the largest score in each column underlined. We will delve into the details of the impacts of different augmentation strategies in the following subsections.

The Impact of Class-Specific Augmentation To explore class-specific augmentation effects, we augmented the datasets using three different modes, augmenting only fake, only real, or both fake and real. In the TALLIP dataset, augmenting with only fake news consistently yields the highest F1 scores with English, Hindi, Indonesian, Vietnamese, and Swahili scores improving by 7.7%,

[9] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

Table 2. F1 scores for BERT-based classifiers on augmented datasets with varying augmentation rates (random sampling) and classes. (Aug.: Augmentation)

Aug.	Aug.	TALLIP						MM-COVID						
Mode	Rate	vi	sw	id	hi	en	All	en	es	fr	hi	it	pt	All
None	0	0.693	0.686	0.675	0.662	0.681	0.799	0.940	0.949	**0.961**	0.936	0.889	0.990	0.947
Only Fake	1	**0.725**	**0.704**	**0.713**	0.668	0.725	**0.801**	**0.951**	**0.964**	0.909	0.932	**0.892**	**0.993**	**0.953**
	3	**0.703**	0.660	**0.717**	**0.706**	**0.758**	**0.801**	**0.950**	0.958	0.940	**0.947**	**0.892**	**0.997**	**0.954**
	5	0.679	0.672	0.691	**0.690**	**0.739**	0.799	0.946	**0.962**	0.932	0.943	0.865	0.987	0.951
Only Real	1	0.634	0.665	0.665	0.647	0.709	0.774	0.940	0.919	0.912	0.938	0.812	0.944	0.929
	3	0.646	**0.690**	0.679	0.659	0.724	0.772	0.930	0.909	0.938	0.925	0.812	0.937	0.919
	5	0.692	0.682	0.676	0.623	0.698	0.773	0.926	0.884	0.921	0.924	0.794	0.948	0.920
Fake + Real	1	0.692	0.668	0.662	0.641	0.688	0.780	0.942	0.955	0.953	0.931	0.829	0.951	0.941
	3	0.665	0.649	0.649	0.614	0.673	0.746	0.942	0.949	0.953	**0.951**	0.829	0.955	0.938
	5	0.644	0.641	0.651	0.618	0.665	0.752	0.931	0.947	0.946	0.943	0.861	0.955	0.936

4.4%, 4.2%, 3.2%, and 1.8% respectively. Augmenting with only real news produced mixed results, with some improvements observed but less consistent. Augmenting with both real and fake news generally led to a reduction in performance compared to the baseline (no augmentation). Similarly, in the MM-COVID dataset, "only fake" class augmentation exhibits the most significant improvements. The "only real" mode results in decreased F1 scores across all languages, indicating that this approach is not effective on this dataset. The "Fake + Real" mode shows mixed results.

The Effects of Augmentation Rates As outlined earlier, we aim to determine the optimal augmentation rate by training the classifiers with datasets having varying numbers of augmented samples. We investigate rates 1, 3, and 5. In both datasets lower augmentation rates generally improve model performance, but as the rate increases, performance tends to decline, suggesting potential overfitting and noise introduction. Generally, rate 1 yields the highest scores, rate 3 shows mixed effects, and rate 5 leads to decreased performance. For instance, in TALLIP, the Vietnamese F1 scores dropped from 0.725 at rate 1 to 0.703 at rate 3 and further decreased to 0.679 at rate 5.

Random vs. Similarity-Based Selection The impact of two subsampling methods on rate 1 augmentation is exhibited in Table 3. The results indicate that similarity-based selection generally leads to higher F1 scores when augmenting both real and fake news samples. However, the difference is less pronounced when only real or fake news is augmented. This limited performance gain may be attributed to the relatively consistent text generated by the Llama 3 model under our specified settings.

5 Limitations and Future Work

One notable issue with large language models (LLMs) is their proneness to produce hallucinations— content that is nonsensical or unfaithful to the source.

Table 3. F1 scores of BERT-based classifiers on rate-one augmented datasets: selection methods comparison. (Aug.: Augmentation, Sel.: Selection, R: Random, S: Similarity)

Aug.	Sel.	TALLIP						MM-COVID						
Mode	Mode	vi	sw	id	hi	en	All	en	es	fr	hi	it	pt	All
None	-	0.693	0.686	0.675	0.662	0.681	0.799	0.940	0.949	0.961	0.936	0.889	0.990	0.947
Only	R	0.725	0.704	0.713	0.668	0.725	0.801	0.951	0.964	0.909	0.932	0.892	0.993	0.953
Fake	S	0.719	0.717	0.715	0.695	0.721	0.789	0.945	0.971	0.925	0.935	0.883	0.993	0.956
Only	R	0.634	0.665	0.665	0.647	0.709	0.774	0.940	0.919	0.912	0.938	0.812	0.944	0.929
Real	S	0.642	0.672	0.667	0.661	0.700	0.773	0.927	0.912	0.912	0.945	0.806	0.944	0.928
Fake +	R	0.692	0.668	0.662	0.641	0.688	0.780	0.942	0.955	0.953	0.931	0.829	0.951	0.941
Real	S	0.706	0.670	0.686	0.653	0.680	0.782	0.952	0.956	0.953	0.949	0.861	0.962	0.940

This can result in irrelevant or unexpected answers when augmenting data, lowering the quality of our augmented datasets and introducing noise. Additionally, similar prompts may yield vastly different outputs, and generated news may undergo changes that alter their intended labels. Grammatical or punctuation errors in the original news may also mislead the model during generation. In our work, we sought to address these issues through iterative testing and experimentation with various prompts, along with randomly inspecting the generated outputs to ensure optimal results. However, this process is time-consuming, and there remains significant room for further improvement. Another limitation is the substantial computational resources required by LLMs. These resources may not always be available, necessitating adjustments to parameters. For instance, shortening the generated output's sequence length can improve generation speed but may also result in cut-off or low-quality augmented data, particularly for longer news articles.

Future work could involve experimenting with different LLMs and settings to improve performance and reduce hallucinations, leading to higher-quality augmented datasets. Employing more advanced prompting techniques can help minimize irrelevance and ensure that the generated content aligns with the intended labels, improving the training data quality. This enhancement will lead to more robust detection models that effectively combat fake news in a timely manner. Additionally, leveraging multilingual models offers noteworthy benefits, as they can capture the subtle nuances of different languages, unlike translation-based approaches that may lose valuable language-specific features.

6 Conclusion

In this study, we addressed the challenge of limited labeled datasets for fake news detection by augmenting the datasets using the capabilities of large language models (LLMs), specifically the open-source Llama 3 model. Our approach involved experimenting with various Llama 3-based augmentation strategies to identify the most effective settings for enhancing the dataset while minimizing noise and the risk of overfitting. We explored various augmentation rates (1, 3,

and 5), the impact of random vs. similarity-based subsampling, and the effects of augmenting specific classes (only fake, only real, or both).

Our results demonstrated the effectiveness of data augmentation in training a more robust model and improving performance compared to the baseline, which did not involve data augmentation. However, it is important to note that excessive augmentation in most cases led to negative effects, resulting in reduced performance or inconsistent outcomes. Our findings highlighted that the "only fake" augmentation mode consistently outperforms the others, and lower augmentation rates generally provide the best balance between introducing useful variability and avoiding overfitting and noise. Notably, augmenting only the fake class in the TALLIP dataset significantly boosted F1 scores, with increases of 7.7% for English and 4.4% for Hindi. In the MM-COVID dataset, this approach resulted in a 1.1% improvement for both languages. Conversely, higher augmentation rates tend to degrade performance, suggesting a threshold beyond which augmentation becomes detrimental to the model's performance. Additionally, our comparison between random and similarity-based sampling showed that similarity-based selection often yields higher F1 scores, though the difference is less notable when augmenting only one class. Overall, employing LLMs for augmenting fake news datasets demonstrated significant potential, highlighting the importance of leveraging state-of-the-art methods to address the challenges of limited labeled data and enhance the performance of fake news detection models.

References

1. Ahlbäck, E., Dougly, M.: Can large language models enhance fake news detection?: Improving fake news detection with data augmentation (2023)
2. Aïmeur, E., Amri, S., Brassard, G.: Fake news, disinformation and misinformation in social media: a review. Social Network Analysis and Mining **13**(1), 30 (2023)
3. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of economic perspectives **31**(2), 211–236 (2017)
4. 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, pp. 2537–2542 (2020)
5. Ashraf, N., Butt, S., Sidorov, G., Gelbukh, A.F.: Cic at checkthat! 2021: Fake news detection using machine learning and data augmentation. In: CLEF (Working Notes), pp. 446–454 (2021)
6. Bayer, M., Kaufhold, M.A., Reuter, C.: A survey on data augmentation for text classification. ACM Computing Surveys **55**(7), 1–39 (2022)
7. Chalehchaleh, R., Farahbakhsh, R., Crespi, N.: Multilingual fake news detection: A study on various models and training scenarios. In: Intelligent Systems Conference, pp. 73–89. Springer (2024)
8. Chiang, W.L., Li, Z., Lin, Z., Sheng, Y., Wu, Z., Zhang, H., Zheng, L., Zhuang, S., Zhuang, Y., Gonzalez, J.E., Stoica, I., Xing, E.P.: Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality (2023). URL https://lmsys.org/blog/2023-03-30-vicuna/
9. De, A., Bandyopadhyay, D., Gain, B., Ekbal, A.: A transformer-based approach to multilingual fake news detection in low-resource languages. Transactions on Asian and Low-Resource Language Information Processing **21**(1), 1–20 (2021)

10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
11. Ding, B., Qin, C., Zhao, R., Luo, T., Li, X., Chen, G., Xia, W., Hu, J., Luu, A.T., Joty, S.: Data augmentation using llms: Data perspectives, learning paradigms and challenges. arXiv preprint arXiv:2403.02990 (2024)
12. D'Ulizia, A., Caschera, M.C., Ferri, F., Grifoni, P.: Fake news detection: a survey of evaluation datasets. PeerJ Computer Science **7**, e518 (2021)
13. Hamed, S.K., Ab Aziz, M.J., Yaakub, M.R.: A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion. Heliyon (2023)
14. Hua, J., Cui, X., Li, X., Tang, K., Zhu, P.: Multimodal fake news detection through data augmentation-based contrastive learning. Applied Soft Computing **136**, 110,125 (2023)
15. Júnior, W.O., da Cruz, M.S., Wyzykowski, A.B.V., de Jesus, A.B.: The use of data augmentation as a technique for improving neural network accuracy in detecting fake news about covid-19. arXiv preprint arXiv:2205.00452 (2022)
16. Kapusta, J., Držík, D., Šteflovič, K., Nagy, K.S.: Text data augmentation techniques for word embeddings in fake news classification. IEEE Access **12**, 31,538–31,550 (2024)
17. Keya, A.J., Wadud, M.A.H., Mridha, M., Alatiyyah, M., Hamid, M.A.: Augfake-bert: handling imbalance through augmentation of fake news using bert to enhance the performance of fake news classification. Applied Sciences **12**(17), 8398 (2022)
18. Li, Y., Jiang, B., Shu, K., Liu, H.: Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation. arXiv preprint arXiv:2011.04088 (2020)
19. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
20. Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., Gao, J.: Large language models: A survey. arXiv preprint arXiv:2402.06196 (2024)
21. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
22. Shanahan, M.: Talking about large language models. Communications of the ACM **67**(2), 68–79 (2024)
23. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
24. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al.: Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288 (2023)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
26. Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., Hu, X.: Harnessing the power of llms in practice: A survey on chatgpt and beyond. ACM Transactions on Knowledge Discovery from Data **18**(6), 1–32 (2024)
27. Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al.: A survey of large language models. arXiv preprint arXiv:2303.18223 (2023)