Artificial intelligence systems for Fake News detection on the Internet

Alexandru Ciobanu
AI Multimedia Lab
National University of Science and Technology POLITEHNICA Bucharest
Bucharest, Romania
alexandru.ciobanu@stud.etti.upb.ro

Bogdan Ionescu
AI Multimedia Lab
National University of Science and Technology POLITEHNICA Bucharest
Bucharest, Romania
bogdan.ionescu@upb.ro

***Abstract*—This research explores the application of artificial intelligence in detecting fake news, addressing challenges of misinformation across multiple languages and domains. A novel multilingual dataset, PolyglotFakeFacts, is developed to improve model accuracy and robustness. We benchmark state-of-the-art models on this dataset, contributing practical advancements for effective fake news detection.**

***Index Terms*—fake news detection, multilingual dataset, natural language processing, transformer models**

I. Introduction

The doctoral research focuses on using artificial intelligence to identify fake news, a pressing issue in today's digital information landscape where misinformation spreads rapidly. The purpose of this study is to develop models capable of detecting fake news within sensitive areas such as politics, security, and geopolitics. Given the global impact of misinformation, this research addresses multilingual detection challenges by creating a unique dataset that includes diverse languages and domains.

II. Stages of research and achieved results

A. Evaluation of the research landscape

To establish a strong foundation, an extensive review of current research was conducted, analyzing 78 articles published between 2020 and 2022. This analysis identified state-of-the-art methods, including Transformer-based models like BERT [4] and GPT-3, as well as multimodal and adversarial approaches. It also highlighted widely-used datasets, such as LIAR [1], ISOT [2], and FakeNewsNet [3], and underscored the need for multilingual datasets to support model robustness across linguistic environments.

B. PolyglotFakeFacts dataset

Recognizing limitations in existing resources, we developed a new multilingual dataset, PolyglotFakeFacts, to enable thorough analysis and training of fake news detection models. This dataset includes both fake and real news articles across domains like politics, social issues, economics, military, and geostrategy. It contains 4912 fake and an equivalent number of real articles, with content in 18 languages. The structure includes fields such as:

TABLE I: PolyglotFakeFacts Dataset Structure

Column	Description
Gathering Date	Date of data collection
News Date	Original publication date
URL	Link to the original article
Source Name	Domain or website
Language	Language of the text
Keywords	Key topics or themes
News Headline	Article title
News Original Text	Full text in the original language
English Translation	English-translated text
Label	Fake or non-fake designation

The dataset was carefully curated, incorporating content verified by the EuvsDisinfo project under the European External Action Service. This resource forms a comprehensive base for analyzing fake news across languages.

C. Testing and benchmarking classification models

We tested three specific models on the PolyglotFakeFacts dataset—BERT [4], LED [5], and RoBERTa [6]. These models were chosen for their diverse architectures and unique strengths in processing text for complex NLP tasks, particularly those requiring contextual understanding and robustness across multilingual data.

- **BERT** [4]: Bidirectional Encoder Representations from Transformers (BERT) was selected for its ability to capture bidirectional context, making it suitable for understanding the nuanced patterns in fake news. BERT's architecture excels in capturing both local and global context in text, which is essential when differentiating genuine content from disinformation.
- **LED (Longformer Encoder-Decoder)** [5]: LED was chosen for its optimization in handling lengthy documents, a common trait in detailed news articles. Unlike standard models that face challenges with extended text sequences, LED uses a sparse attention mechanism to efficiently process long documents, allowing it to retain context across entire articles. This attribute is particularly valuable in fake news detection, as fake news articles often contain extended arguments or complex narratives.

1st Workshop on Artificial Intelligence for Multimedia, AI Multimedia Lab, CAMPUS Research Institute, POLITEHNICA Bucharest, Bucharest, Romania, 8 November, 2024

- **RoBERTa** [6]: Robustly optimized BERT approach (RoBERTa) builds upon BERT's architecture with adjustments for improved performance, such as longer training on more data and removing the next sentence prediction task. This makes RoBERTa highly effective at detecting subtle contextual clues in text, providing high accuracy in distinguishing between real and fake news. RoBERTa's optimized architecture has shown state-of-the-art performance on various NLP benchmarks, making it a valuable choice for fake news classification.

The evaluation of these models involved key performance metrics, as shown below:

TABLE II: Model Benchmark Results on PolyglotFakeFacts

Model	Accuracy	Precision	Recall	F1-Score
BERT	87.5%	88.0%	86.0%	87.0%
LED	85.2%	85.5%	84.0%	84.7%
RoBERTa	89.0%	89.5%	88.0%	88.7%

These results offer valuable insights into the relative strengths of each model. BERT and RoBERTa displayed strong performances, with RoBERTa slightly outperforming BERT due to its robust pre-training optimizations. LED, while slightly lower in overall metrics, provided substantial benefits when dealing with longer texts, highlighting the importance of model selection based on document length and complexity. These insights will guide future developments, focusing on optimizing models for fake news detection across diverse linguistic and structural contexts.

III. Future directions and practical steps

A. Enhancing dataset diversity

Future work will expand PolyglotFakeFacts to include more languages and sources. Preprocessing methods like paraphrasing and translation will improve diversity, enhancing model robustness.

B. Hybrid and advanced deep learning models

Hybrid models combining traditional NLP with advanced networks will be explored, alongside ensemble methods to maximize detection performance. Newer models, like DeBERTa [7] and GPT-4 [8], will be benchmarked for linguistic versatility

.

C. Explainability techniques

To validate model predictions, LIME [9] and SHAP [10] will provide transparency, clarifying the features that influence detection, essential for user trust.

D. Fine-tuning for multilingual detection

Fine-tuning across languages and domains will optimize the model's adaptability, ensuring it performs effectively in varied contexts.

IV. Conclusion

This research emphasizes practical methods for developing an AI system capable of detecting fake news across languages and domains. Focusing on dataset diversity, model enhancement, and explainability, it provides a robust foundation to combat misinformation globally.

References

[1] W. Y. Wang, "Liar, liar pants on fire: A new benchmark dataset for fake news detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017.

[2] D. Ahmed H., L. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Security and Privacy*, vol. 1, no. 1, pp. 1-9, 2018.

[3] H. Shu, S. Wang, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for fake news research," in *Proceedings of the 10th ACM International Conference on Web Search and Data Mining (WSDM)*, 2017.

[4] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019.

[5] A. Beltagy, M. Peters, and A. Cohan, "Longformer: The long-document transformer," *arXiv preprint arXiv:2004.05150*, 2020.

[6] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[7] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with disentangled attention," in *International Conference on Learning Representations (ICLR)*, 2021.

[8] OpenAI, "GPT-4 Technical Report," *OpenAI*, 2023. Available: https://openai.com/research/gpt-4

[9] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.

[10] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.