1 Introduction

In today's digital era, the widespread use of the internet and social media has dramatically transformed the way information is created, disseminated, and consumed. While these advancements offer unprecedented access to knowledge and communication, they have also facilitated the rapid spread of misinformation, particularly fake news. Fake news can originate from disinformation (intentionally fabricated content for specific purposes) or evolve as misinformation (false content shared unknowingly by individuals who believe it to be true). Psychological factors, such as confirmation bias, further exacerbate the issue; people tend to accept and share information that aligns with their personal beliefs, regardless of its factual accuracy. Behavioral studies suggest that exposure to fake news increases the likelihood of belief in the content, and even when individuals recognize the content as false, they may still share it to affirm social identity, signal political alignment, or elicit emotional reactions from their network [1] .

An MIT study that tracked 126,000 stories revealed an alarming trend, fake news stories are approximately 70% more likely to be retweeted on Twitter than factual stories and reach audiences six times faster than the truth [2].The proliferation of fake news has severe implications for societal trust and democratic institutions. For example, recent surveys indicate that two-thirds of EU citizens encounter fake news at least once a week, and over 80% consider it a significant issue for their country and democracy in general [3].

As the challenge of combating fake news becomes increasingly urgent, researchers have increasingly relied on artificial intelligence (AI) solutions to address this complex issue. Initially, traditional machine learning models were leveraged to detect fake news, utilizing various natural language processing (NLP) techniques to preprocess text data. This preprocessing often included tasks like tokenization, lemmatization, and the removal of stop words. Following this stage, algorithms such as Decision Trees, Support Vector Machines (SVMs), or gradient boosting models like XGBoost were commonly employed to classify news content based on linguistic patterns, sentiment markers, and other features that might indicate falsified information. These models provided some initial successes but often struggled with the nuances and evolving language seen in modern misinformation campaigns.

However, recent breakthroughs in Large Language Models (LLMs), like OpenAI's GPT models, have transformed the domain of fake news detection. Large Language Models, constructed with billions of parameters and trained on extensive datasets, demonstrate proficiency in comprehending linguistic context and producing human-like prose. These characteristics allow LLMs to transcend mere feature extraction and concentrate on the semantic and contextual understanding of news content, which is crucial for recognizing the nuanced indicators frequently seen in misleading information. LLMs have shown considerable promise, improving detection accuracy and robustness in ways traditional ML models could not achieve alone.

In our paper, we aim to investigate two key questions that remain central to the effectiveness of AI-driven fake news detection:

1. Are LLMs more accurate in detecting fake news than traditional ML algorithms? This question seeks to assess the performance differences between traditional ML

algorithms and modern LLMs in identifying fake news. By comparing metrics such as accuracy, precision, recall, and F1-score across various models, we aim to quantify whether the contextual understanding afforded by LLMs translates to better detection rates.

2. Do hybrid models that combine Large Language Models (LLMs) with traditional machine learning models or Small Language Models (SMLs) offer improvements in fake news detection? Hybrid models, which combine the strengths of LLMs with more lightweight or interpretable models, represent an intriguing area of research. These combinations could balance the computational intensity of LLMs with the efficiency and interpretability of smaller models, potentially enhancing overall model performance and providing faster, more resource-efficient solutions.

2 Related Work

2.1 Fake News Detection Using Machine Learning Approaches

In their paper, Khanam et al.[4] examines the use of machine learning (ML) approaches to detect fake news on social media, specifically focusing on the LIAR dataset. The study utilizes six ML algorithms—XGBoost, Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machine (SVM)—to analyze political news and classify it as fake or real. For feature extraction and processing they used Natural Language Processing (NLP) techniques, including tokenization, feature extraction with TF-IDF, and parts of speech tagging, are applied to refine data and improve model input quality. XGBoost demonstrated the highest accuracy (75%), followed by SVM and Random Forest, both achieving around 73% accuracy.

2.2 Large Language Model Based Fake News Detection

Aman [5] uses the LLaMA language model to detect deception in the form of text, images, and videos produced by deepfake technology. Through "self-instruction" and "alignment" techniques, it focuses on task-specific fine-tuning, enabling the model to detect bogus news efficiently even with constrained computational resources. To maximize model performance without requiring a lot of technology, key techniques include mixed-precision quantization and parameter-efficient tweaking. Despite encouraging outcomes, managing computing costs and identifying realistic deepfakes continue to be difficult tasks.

2.3 Exploring the Role of Large Language Models in Fake News Detection

Beizhe Hu et al [6] examines the potential function of large language models (LLMs), such as GPT-3.5, in the identification of fake news. Although LLMs have sophisticated reasoning capabilities, they remain inferior to fine-tuned small language models (SLMs), like BERT, in the domain of task-specific fake news identification. The authors introduce the Adaptive Rationale Guidance (ARG) network, wherein supervised learning models (SLMs) utilize rationales derived from large language models (LLMs) to enhance accuracy without necessitating direct involvement from LLMs. The ARG network, together with its condensed variant (ARG-D) for cost-sensitive applications, exhibited superior performance compared to the utilization of LLMs or SLMs independently on English and Chinese datasets. This study emphasizes that LLMs can assist SLMs by providing multi-faceted analysis, however obstacles persist in properly integrating these insights for practical implementations.

2.4 Integrating Large Language Models and Machine Learning for Fake News Detection

Teo et al. [7] explores how to enhance false news identification by combining conventional machine learning (ML) techniques with Large Language Models (LLMs), particularly ChatGPT-3.5 and Bard. The study shows that hybrid models perform better, especially when XGBoost is used with LLM predictions. The XGBoost hybrid model achieved 96.39% accuracy, 97.04% precision, 98.17% recall, and a 97.6% F1 score. The hybrid strategy improves typical ML models' capacity to discern between authentic and fraudulent news by utilizing LLMs' sophisticated linguistic understanding.

2.5 Fake News Detection with Large Language Models on the LIAR Dataset

Boissonneault et al. [8] evaluate the efficacy of large language models (LLMs), namely ChatGPT and Google Gemini, in detecting false information. The authors utilized the LIAR dataset to evaluate the models' performance, discovering elevated accuracy, precision, recall, F1 scores, and AUC-ROC values, with Google Gemini marginally surpassing ChatGPT. This study emphasizes the capacity of LLMs to enhance fact-checking systems, assisting media and social platforms in mitigating misinformation. Primary problems encompass managing hostile inputs and biases throughout training data. Research indicates that including LLMs into content verification may improve the accuracy of automated fact-checking; nevertheless, additional refinement is necessary to mitigate restrictions, including computational resource requirements.

2.6 Comparative Evaluation of News Verifiers: ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard

Caramancion [9] compares ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking . The experiment was conducted in a controlled manner, evaluating OpenAI's ChatGPT (versions 3.5 and 4.0), Google's Bard, and Microsoft's Bing AI with 100 verified news items sourced from independent agencies. The results indicated an average accuracy score of 65.25%, with GPT-4.0 achieving the highest score of 71%, signifying modest ability in identifying bogus information. Not withstanding these gains, human fact-checkers continue to outperform due to their sophisticated comprehension of context, underscoring the limitations of LLMs in precisely analyzing intricate news data.

3 Dataset

3.1 LIAR

Selected over a decade from PolitiFact.com, the publicly available LIAR dataset, developed by Wang [10], consists of 12.8K brief statements classified for truthfulness. Every LIAR record features not just the statement but also a thorough analysis, source references, and metadata including speaker job title, party affiliation, and historical truthfulness. With columns for the statement ID, truthfulness label (e.g., true, largely true, false), statement text, subject, speaker details, and context, the dataset is TSV (tab-separated values).

3.2 FakeNewsNet

In addition to the LIAR dataset, this study utilizes FakeNewsNet which is a notable dataset developed by Shu et al.[11], concentrating on news content and its social context on platforms such as Twitter. Derived from PolitiFact and GossipCop, it comprises CSV files that classify news as authentic or fabricated, along with connections to tweets disseminating the news. In compliance with privacy regulations, FakeNewsNet employs a tool, FakeNewsTracker, to assist academics in analyzing and visualizing the spread of false news on social media while adhering to Twitter's principles. This dataset facilitates research on the social dynamics and attributes of fake vs authentic news, rendering it essential for investigations into misinformation across many platforms.

References

[1] Bitesize, B.: What is misinformation? Accessed: 2024-11-12 (2024). https://www.bbc.co.uk/bitesize/articles/zcr8r2p

[2] Friggeri, A., Garcia, E.E., González, L.A.: Study: False news travels faster than true stories on twitter. MIT News (2018). Accessed: 2024-11-12

[3] Europe, C.: Dealing with Propaganda, Misinformation and Fake News. Accessed: 2024-11-12 (2024). https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/dealing-with-propaganda-misinformation-and-fake-news

[4] Khanam, Z., Alwasel, B.N., Sirafi, H., Rashid, M.: Fake news detection using machine learning approaches. IOP Conference Series: Materials Science and Engineering **1099**, 012040 (2021) https://doi.org/10.1088/1757-899X/1099/1/012040 . International Conference on Applied Scientific Computational Intelligence using Data Science (ASCI 2020), 22nd-23rd December 2020, Jaipur, India

[5] Aman, M.: Large language model based fake news detection. In: Soft Computing and Intelligent Systems: Theory and Applications (EUSPN 2023). Elsevier B.V.,

Almaty, Kazakhstan (2023). Kazakh-British Technology University. Available online at www.sciencedirect.com

[6] Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., Qi, P.: Bad actor, good advisor: Exploring the role of large language models in fake news detection. In: Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24), Vancouver, BC, Canada (2024). Association for the Advancement of Artificial Intelligence. CAS Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences; National University of Singapore

[7] Teo, T.W., Chua, H.N., Jasser, M.B., Wong, R.T.K.: Integrating large language models and machine learning for fake news detection. In: 2024 20th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, Langkawi, Malaysia (2024). Department of Computing and Information Systems, Sunway University

[8] Boissonneault, D., Hensen, E.: Fake news detection with large language models on the liar dataset. Research Article (2024) https://doi.org/10.21203/rs.3.rs-4465815/v1 . Creative Commons Attribution 4.0 International License

[9] Caramancion, K.M.: News verifiers showdown: A comparative performance evaluation of chatgpt 3.5, chatgpt 4.0, bing ai, and bard in news fact-checking. University of Wisconsin–Stout Mathematics, Statistics, and Computer Science Department (2024). Available at: www.kevincaramancion.com

[10] Wang, W.Y.: "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), Short Papers, pp. 422–426. Association for Computational Linguistics, Vancouver, BC, Canada (2017)

[11] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:1809.01286 (2018)