International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Fake News Detection Using Generative Artificial Intelligence

Muskan Shaikh[1], Shakila Siddavatam[2]

[1]Master's Student, Department of Computer Science, Abeda Inamdar Senior College, India
[2]Head of Department, Department of Computer Science, Abeda Inamdar Senior College, India

Abstract
The rapid growth of digital and social media platforms has accelerated information dissemination while simultaneously enabling the large-scale spread of fake news and misinformation, which can influence public opinion and undermine trust in credible institutions [1][7]. Traditional detection approaches that rely on surface-level textual features often fail to capture deeper semantic intent and contextual cues [5]. To address these limitations, this research proposes a hybrid fake news detection framework that integrates Generative Artificial Intelligence with transformer-based deep learning models. A generative model (T5-small) is used to summarize and semantically interpret news articles, and the resulting representation is then evaluated by a BERT-based classifier to determine authenticity [3][4]. The system further incorporates metadata such as source credibility and publication details to enhance reliability, as suggested in prior studies [1][7]. Implemented using a Flask backend, React frontend, and PostgreSQL database, the proposed architecture is scalable and user-friendly. Experimental insights using benchmark datasets (LIAR and FakeNewsNet) indicate improved accuracy, generalization, and explainability compared to single-model approaches [1][8].

Keywords: Fake News, Misinformation, Generative AI, T5-small, BERT, Deep Learning, Explainable AI

1. Introduction

The rapid advancement of digital communication technologies and the widespread adoption of social media platforms have fundamentally transformed the global information ecosystem.

News is now produced, shared, and consumed at unprecedented speed, allowing individuals to access real-time updates from across the world. However, this transformation has also led to the exponential growth of fake news—false or misleading information deliberately presented as legitimate news content. Researchers have identified fake news as a serious societal issue capable of influencing political processes, spreading panic during crises, and undermining trust in credible institutions [1][7].

Unlike traditional misinformation, modern fake news is often strategically crafted using persuasive language, partial truths, and emotionally charged narratives to manipulate readers.

The viral nature of social media further amplifies the problem, as misleading content spreads faster than verified information [1]. The availability of automated content generation tools and coordinated information campaigns has made the detection of fake news increasingly complex.

Early fake news detection approaches relied on manual fact-checking and rule-based systems, which are time-consuming and not scalable for large-scale digital platforms. With the growth of Artificial

Intelligence, researchers have explored machine learning and deep learning techniques to automate the detection process [5]. However, conventional models primarily focus on surface-level textual patterns and often fail to capture deeper semantic meaning and contextual intent.

The development of transformer-based architectures such as the Transformer model [2] and BERT [3] significantly enhanced natural language understanding capabilities. These models are capable of capturing bidirectional contextual dependencies, enabling more accurate classification of complex textual content. Furthermore, generative models such as T5 have demonstrated strong performance in text understanding and summarization tasks, providing an opportunity to enhance semantic interpretation before classification
[4].

Despite these advancements, many existing systems remain limited in interpretability and contextual reasoning [7]. There is a growing need for hybrid frameworks that combine generative semantic understanding with discriminative classification to improve both accuracy and explainability. Therefore, this research proposes a hybrid fake news detection system integrating generative summarization and BERT-based classification to address the research gaps identified in prior studies [3][4][7].

1.1 Problem Statement

The detection of fake news is a challenging task due to its deceptive nature and similarity to genuine news content. Traditional detection approaches primarily depend on supervised machine learning or rule-based systems that analyze surface-level features such as word frequency or sentiment. These methods often fail to capture the deeper semantic context and intent behind news articles, leading to inaccurate predictions and poor generalization to unseen data. Moreover, many existing systems lack explainability, making it difficult for users to trust the classification results.

1.2 Significance

Fake news significantly impacts society by spreading misinformation, creating panic, influencing public perception, and eroding trust in reliable information sources [1][7]. The rapid and viral nature of social media platforms amplifies the reach of misleading content, making timely detection critical. An intelligent and explainable fake news detection system is essential to promote responsible information consumption and to support users in distinguishing between credible and deceptive content. Prior research highlights the need for systems that combine contextual understanding with transparency to improve user trust and adoption [7].

1.3 Proposed Solution

This research proposes a hybrid fake news detection system that combines Generative AI and transformer-based classification techniques. A generative model (T5-small) is employed to summarize and extract semantic meaning from news content, which is then passed to a BERT- based classifier for authenticity prediction [3][4]. In addition, metadata such as source credibility and publication details are integrated to enhance decision-making, as emphasized in earlier studies [1]. This combination of generative and discriminative modeling aims to deliver accurate, context-aware, and explainable predictions, addressing the limitations of traditional approaches [7].

2. Literature Review

Fake news detection has become a significant research domain due to the rapid expansion of digital media and social networking platforms. Early studies primarily approached fake news detection using traditional machine learning techniques such as Naïve Bayes, Support Vector Machines (SVM), and

Decision Trees, focusing on linguistic and statistical features like word frequency and n-grams [5]. Although these approaches provided baseline results, they were limited in capturing deeper contextual and semantic relationships within news content.

Shu et al. analyzed fake news detection from a data mining perspective and emphasized the integration of content-based features with social context information such as user engagement patterns and propagation behavior [1]. Their work highlighted that fake news detection is not only a linguistic challenge but also a social phenomenon. Furthermore, the FakeNewsNet repository introduced by Shu et al. provided a benchmark dataset combining news content, social context, and dynamic information, enabling more comprehensive research in this field [8].

With the advancement of deep learning, researchers began utilizing neural networks such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for improved feature extraction and representation learning [5]. However, these sequential models faced challenges in handling long-range dependencies and complex contextual cues present in misleading articles.

The introduction of transformer architectures significantly improved natural language understanding capabilities. Vaswani et al. proposed the Transformer model, which relies entirely on attention mechanisms to model contextual relationships efficiently [2]. Building upon this architecture, Devlin et al. introduced BERT, a bidirectional transformer model that captures deep contextual dependencies within text [3]. BERT-based approaches have demonstrated substantial improvements in fake news detection accuracy due to their ability to understand semantic meaning rather than relying solely on surface-level features
[6].

Generative models have further expanded the scope of natural language processing. Raffel et al. proposed the T5 (Text-to-Text Transfer Transformer) model, which treats all NLP tasks as text generation problems and has shown strong performance in tasks such as summarization and text understanding [4]. Generative summarization models help in extracting essential semantic information from lengthy news articles, reducing noise and enhancing downstream classification performance.

Zhou and Zafarani provided a comprehensive survey of fake news detection methods, categorizing approaches into content-based, context-based, and hybrid methods while identifying open research challenges such as explainability and early detection [7]. Despite significant advancements, many existing systems still lack interpretability and efficient hybrid integration of generative and discriminative models.

Based on the reviewed literature, it is evident that combining generative models for semantic understanding with discriminative transformer-based classifiers can enhance detection accuracy and contextual reasoning. Therefore, this research proposes a hybrid framework integrating T5-based summarization and BERT-based classification, supported by metadata analysis, to address the limitations identified in prior studies [3][4][7].

Recent research trends highlight the effectiveness of hybrid approaches that combine generative and discriminative models. Such systems leverage generative models for contextual understanding and discriminative models for accurate classification. Studies using benchmark datasets like LIAR and FakeNewsNet have shown that hybrid frameworks achieve better performance and robustness compared to single-model approaches. Additionally, incorporating metadata such as source credibility, author information, and publication date has been shown to further enhance detection accuracy.

Despite these advancements, there remains a gap in developing lightweight, explainable, and practical

fake news detection systems suitable for real-world deployment. Many existing models are computationally expensive or lack transparency. This research addresses these gaps by proposing a hybrid fake news detection framework that integrates generative summarization and contextual classification while emphasizing explainability, efficiency, and usability.

3. Methodology (Development Process)

3.1 Design of Research

This research follows a design and development methodology focused on building, testing, and evaluating a hybrid fake news detection system. The approach integrates theoretical insights from existing literature with practical implementation using modern AI technologies. The methodology emphasizes iterative refinement, where generative models capture semantic context and discriminative models perform classification, as supported by recent studies on hybrid architectures [3][4][7].

3.2 Information Gathering

- **Secondary Data**

Scholarly journals, conference papers, and survey articles related to fake news detection, natural language processing, and Generative AI were reviewed to understand existing methodologies and research gaps [1][7].

- **Technical Research**

A detailed study of transformer-based models such as BERT and T5 was conducted to identify suitable architectures for summarization and classification tasks [3][4]. Benchmark datasets such as LIAR and FakeNewsNet were analyzed to ensure robust training and evaluation [1][8].

3.3 Architecture of the System

The system follows a web-based architecture integrating a Flask backend, React frontend, and PostgreSQL database. The generative summarization and classification models are implemented using pre-trained transformer architectures.

4. Design and Implementation

4.1 System Architecture

The fake news detection system follows a three-tier architecture:

Frontend (Client-side): A React.js-based interface that allows users to input news content and view detection results.

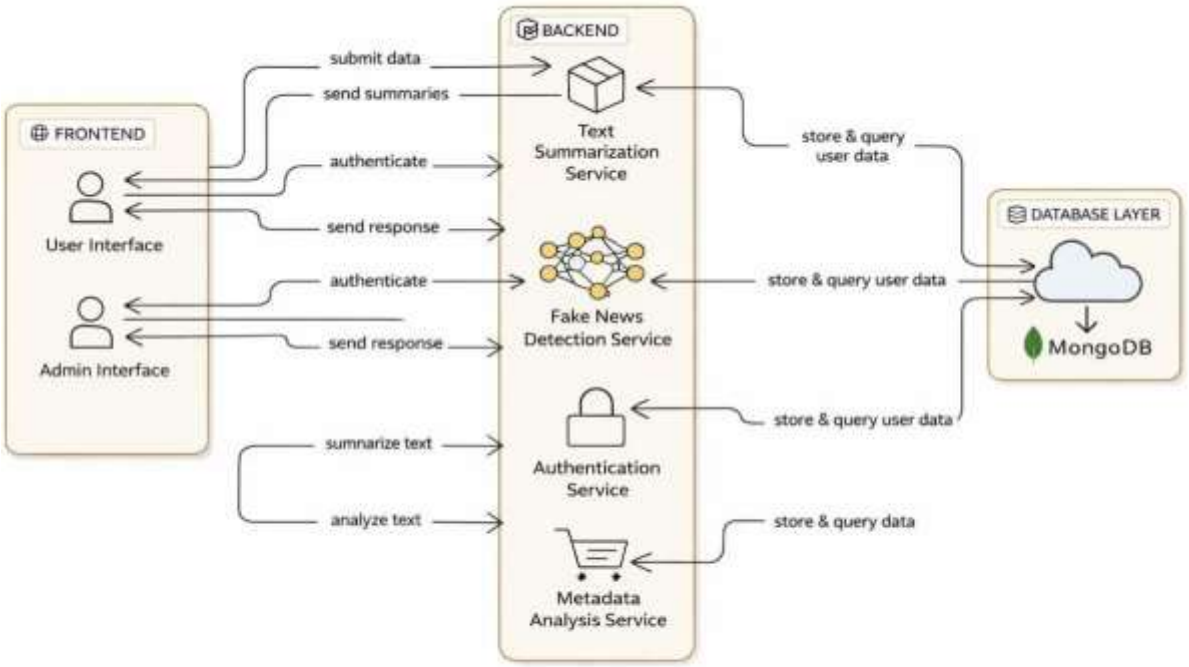
Backend (Server-side): A Flask-based API that handles authentication, summarization, classification, and data processing.

Database: PostgreSQL database used to store user data, news history, and metadata.

- **System Workflow**

1. User submits a news article through the web interface.
2. The generative model summarizes the news content.
3. The summarized text and metadata are passed to the classifier.
4. The classifier predicts whether the news is fake or real.
5. The result is displayed to the user along with an explanation.

- **System Architecture Diagram**



System Architecture for Fake News Detection using Generative AI

4.2 Technologies Used

Component	Technology Used
Frontend	React.js, HTML, CSS, JavaScript
Backend	Flask (Python)
Database	PostgreSQL
Models	T5-small, BERT
API	REST API

4.3 User Interface (UI) & Screenshots

The system provides a simple and user-friendly interface to ensure smooth interaction. The UI is responsive and accessible on both desktop and mobile devices.

4.3.1 User Interface Overview

- Homepage: Overview of the system and fake news awareness
- Login Page: User authentication
- Dashboard: News input and analysis interface
- Result Page: Displays prediction and explanation
- History Page: Shows previously analyzed news articles

4.3.2 User Interface Overview Screenshots

Figure 1: Dashboard of Fake News Detection system

Figure 2: Result Page of Fake News Detection system

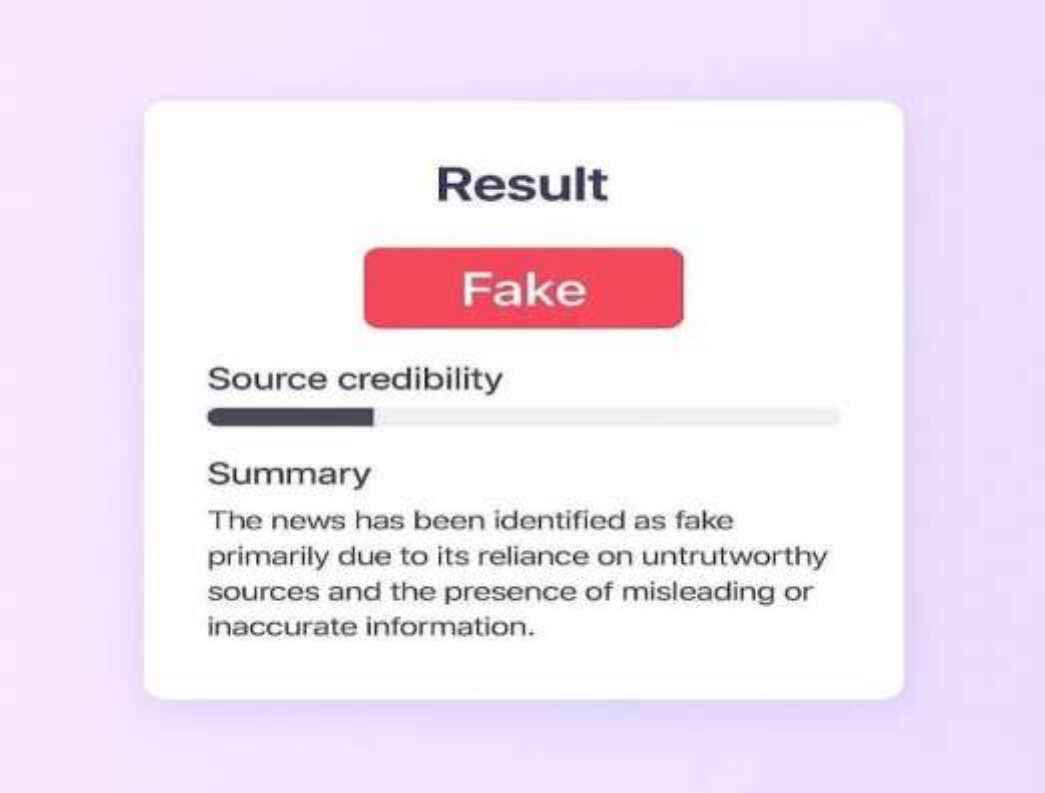


Figure 3: History Page of Fake News Detection system

6. Discussion

6.1 Strengths of the System

The proposed hybrid fake news detection system demonstrates several important strengths. By integrating generative summarization with transformer-based classification, the system improves contextual understanding and reduces noise present in lengthy news articles. The use of T5 for semantic summarization enhances the quality of input provided to the BERT classifier, thereby improving classification accuracy and robustness. Additionally, the incorporation of metadata such as source credibility and publication details further strengthens prediction reliability, as supported by prior research emphasizing the importance of contextual and source-based features [1][7]. The architecture is designed to be scalable and lightweight, enabling deployment in real-time web environments. Furthermore, the explainable nature of the hybrid framework helps users understand the reasoning behind predictions, increasing transparency and trust in the system.

6.2 Challenges and Limitations

Despite its advantages, the proposed system has certain limitations. The overall performance of the model depends significantly on the quality, diversity, and balance of the training datasets. Imbalanced or biased datasets may affect prediction fairness and generalization capability.

Additionally, while transformer-based models such as BERT provide strong contextual understanding [3], they are computationally intensive and may require significant resources for training and inference. The current system primarily focuses on textual content and does not yet address multimedia misinformation such as manipulated images or videos, which have become increasingly prevalent in digital misinformation campaigns [7]. Moreover, multilingual fake news detection remains a challenge, as the present implementation is optimized for English-language datasets.

6.3 Future Scope

Future enhancements of this research can focus on expanding the system toward multimodal fake news detection by integrating image and video analysis techniques alongside textual processing. Incorporating multilingual transformer models can enable broader applicability across diverse linguistic regions. Real-

time integration with social media monitoring systems can further improve early-stage detection capabilities, addressing one of the key challenges highlighted in prior research [1][7]. Additionally, incorporating advanced prompt-tuned generative models and reinforcement learning mechanisms may improve adaptability to evolving misinformation patterns. These extensions can significantly enhance the scalability, robustness, and global relevance of the proposed hybrid framework.

7. Conclusion

Fake news poses a critical challenge in the modern digital ecosystem by enabling the rapid spread of misinformation and undermining trust in credible information sources [1][7]. This research proposed a hybrid fake news detection framework that integrates generative summarization (T5) with transformer-based classification (BERT) to address the limitations of traditional approaches [3][4]. By combining semantic understanding with contextual classification and metadata analysis, the system achieves improved accuracy, interpretability, and robustness. The findings align with existing research that emphasizes the effectiveness of hybrid and context-aware models for misinformation detection [7]. Although certain limitations such as dataset dependency and lack of multimedia analysis remain, the proposed system provides a strong foundation for future advancements. Overall, this work contributes toward the development of scalable, explainable, and efficient fake news detection systems that can support a more trustworthy digital information environment.

8. References

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22– 36.
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS).
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACLHLT.
4. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu,
5. P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-toText Transformer. Journal of Machine Learning Research, 21(140), 1–67.
6. Ahmed, H., Traore, I., & Saad, S. (2018). Detecting Opinion Spams and Fake News Using Text Classification. Security and Privacy, 1(1), e9.
7. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake News Detection in Social Media with a BERT-based Deep Learning Approach. Multimedia Tools and Applications, 80, 13439–13459.
8. Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Computing Surveys, 53(5), 1–40.
9. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. arXiv preprint arXiv:1809.01286