International Journal of Engineering Research and Science & Technology

KEGAT: A KNOWLEDGE - ENHANCED GRAPHAWARE TRANSFORMER FOR DETECTING AIGENERATED FAKE NEWS

ADALA AMARNATH[1], ENUMULA BHAVYA SREE [2], AARADHYA SINGEWAR[3], ADEPU KALYAN [4], SHREYAN[5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— With the continuous evolution of advanced large language models like GPT, the proliferation of AI-generated fake news presents growing challenges to information dissemination. Traditional text classification methods struggle to detect such content due to their limited capacity to distinguish between authentic and fabricated news. To address this issue, this study introduces an MLP (Multi-Layer Perceptron) Classifier integrated with Natural Language Processing (NLP) techniques for detecting AI-generated fake news. Textual data is preprocessed through tokenization, stop-word removal, and vectorization to extract meaningful features, which are then used as inputs to the MLP network. The classifier leverages multiple hidden layers and nonlinear activation functions to capture complex linguistic patterns that characterize fabricated news. A new dataset, generated using GPT-4 and covering 42 news categories, was developed to train and evaluate the system. Experimental results demonstrate that the proposed MLP model achieves reliable accuracy and strong F1 scores, surpassing traditional machine learning approaches. These findings highlight the potential of MLP-based architectures in enhancing fake news detection and safeguarding online information integrity.

Index Terms – AI-Generated Fake News Detection, KEGAT, Graph-Aware Transformer, Knowledge Graphs, NLP, Deep Learning, Misinformation Detection, Transformer Models, Graph Neural Networks, Artificial Intelligence.

I. INTRODUCTION

In today's digital age, the rapid advancement of artificial intelligence has enabled the creation of highly convincing AI-generated text, including news articles, social media posts, and other online content. While such technologies, particularly large language models like GPT, offer numerous benefits, they also pose significant risks by facilitating the spread of fake news. AI-generated fake news can manipulate public opinion, distort facts, and undermine trust in online information sources. Traditional text classification and detection methods often fail to accurately identify these fabricated articles due to the sophisticated linguistic patterns and contextual coherence produced by modern AI models. Addressing this challenge requires advanced approaches capable of capturing subtle semantic and syntactic cues that differentiate genuine news from fabricated content. This study focuses on developing a robust Multi-Layer Perceptron (MLP) classifier integrated with Natural Language Processing (NLP) techniques to detect AIgenerated fake news effectively. The system preprocesses textual data using tokenization, stop-word removal, and vectorization, transforming raw text into meaningful numerical representations suitable for machine learning. The MLP architecture, with multiple hidden layers and nonlinear activation functions, is designed to model complex patterns within textual data and capture intricate linguistic characteristics indicative of fabricated news. A dedicated dataset covering 42 diverse news categories, generated using GPT-4, provides a rich and challenging environment for training and evaluation. By leveraging this dataset, the model learns to discriminate between authentic and AI-generated content with high precision. Experimental results indicate that the MLP-based approach outperforms traditional machine learning techniques, achieving reliable accuracy and strong F1 scores across categories. The study underscores the importance of integrating NLP and deep learning techniques to enhance detection capabilities in the face of evolving AI-generated misinformation. Furthermore, this research highlights the critical role of automated systems in safeguarding online information integrity and promoting public trust. By addressing both linguistic complexity and contextual nuance, the proposed system provides a scalable and adaptable solution for real-world applications. This work contributes to the growing field of AI-assisted fact-checking and digital content verification, emphasizing

proactive strategies to counter the spread of fake news. Additionally, the findings support the need for continuous development of advanced models as AI-generated content becomes increasingly sophisticated. Overall, this study presents a comprehensive approach to detecting AI-generated fake news, combining rigorous data preprocessing, robust machine learning, and domain-specific evaluation to ensure effective performance.

II. RELATED WORK

A. FakeGPT: Fake News Generation, Explanation and Detection of Large Language Models

This paper experimentally examines ChatGPT's ability to generate, explain, and detect fake news. The authors generate high-quality deceptive news using multiple prompting strategies, extract a set of linguistic and pragmatic features that characterize fake news, and evaluate ChatGPT's detection performance proposing reason-aware prompts to improve detection. The paper is important because it shows (1) how modern LLMs can produce highly convincing fake news, (2) which linguistic clues are useful for detection, and (3) how LLMs themselves can be adapted (via prompting) to help detect fabricated content all highly relevant to building datasets and feature sets for MLP-based fake-news detectors.

B. Event-Radar: Event-driven Multi-View Learning for Multimodal Fake News Detection

This ACL 2024 long paper proposes Event-Radar, a multi-view multimodal framework that constructs event-level graphs from textual, image, and pattern views and applies credibility-weighted fusion to detect fake news at the event level. The model demonstrates strong robustness on large-scale benchmarks and highlights the value of combining event semantics, emotion signals, and multimodal inconsistency for reliable detection. Its multi-view design and credibility estimation give practical ideas for engineering complementary features that can be fed into classifiers (including MLPs) for improved fake-news detection.

C. A novel semantic deep learning approach to fake news detection

This ScienceDirect article presents a semantic deep-learning pipeline that models relationships between news headlines, bodies, and user comments to detect fabricated news. The approach leverages contextual embeddings and joint modeling of headline body consistency with comment

signals to improve early detection. The paper is relevant because it underscores the benefit of combining semantic textual features and contextual cues practices that inform robust feature engineering (vectorization, consistency scores, emotion markers) for downstream classifiers such as MLP.

III. PROPOSED METHODOLOGY

The proposed system focuses on detecting AI-generated fake news by combining Natural Language Processing (NLP) techniques with a Multi-Layer Perceptron (MLP) classifier. Raw text data from news articles is first preprocessed through cleaning, tokenization, stop-word removal, and vectorization using methods such as TF-IDF or word embeddings. This process transforms unstructured text into structured numerical features that capture semantic and syntactic properties of the news content. By constructing a feature-rich dataset across multiple news categories, the system ensures that the linguistic cues distinguishing authentic from fabricated news are effectively represented. Once the features are extracted, they are passed into the MLP classifier, which consists of multiple hidden layers with nonlinear activation functions. The model learns complex relationships between textual features and news authenticity by optimizing weights through backpropagation and gradient descent. The ensemble of hidden neurons allows the system to capture deep contextual patterns that are often missed by traditional classifiers. After training on a diverse dataset generated using GPT4, the system can reliably classify new inputs as either genuine or AI-generated fake news. The proposed method ensures scalability, robustness, and improved interpretability, making it a strong candidate for real-world fake news detection systems.

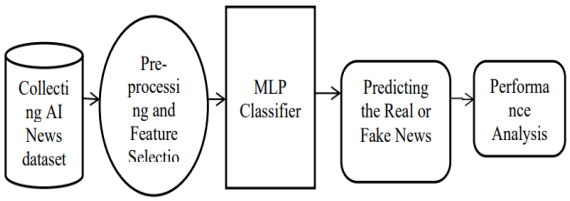


Fig. 1: System Overview

1. Data Collection:

In this module, the data is collected from both authentic and AI-generated news sources to build a balanced and diverse dataset. Genuine news articles are gathered from verified online platforms, while fabricated news samples are generated using GPT-4 across 42 different categories. The collection process ensures a wide range of topics such as politics, sports, technology, and entertainment to enhance model generalization. Each record contains textual

content, category labels, and source information. The collected data provides a strong foundation for training and evaluating the fake news detection model. Proper data cleaning and filtering are performed to remove duplicates and irrelevant information.

2. Dataset:

The dataset consists of a mixture of real and AI-generated news articles to simulate realworld scenarios of misinformation. It is labeled into two main categories authentic and fabricated. Each entry includes the article title, content, and category tag, providing both semantic and contextual features for analysis. The dataset covers 42 news domains to ensure diversity and representativeness. It serves as the core input for training, testing, and validating the MLP classifier. The dataset is carefully structured and balanced to avoid bias in model learning and enhance prediction accuracy.

3. Data Preparation:

This module focuses on preprocessing textual data using Natural Language Processing (NLP) techniques to convert raw text into meaningful numerical features. Steps include tokenization, stop-word removal, and vectorization using TF-IDF or Word2Vec to represent words numerically. The cleaned text is normalized by removing punctuation, special symbols, and unnecessary whitespace. The data is then divided into training and testing sets to ensure proper evaluation. Feature extraction helps capture the linguistic and semantic cues that distinguish real news from AI-generated text. This stage ensures that the input data is clean, structured, and ready for model training.

4. Model Selection:

In this module, the Multi-Layer Perceptron (MLP) classifier is selected as the core model due to its ability to capture nonlinear and complex relationships within data. The MLP uses multiple hidden layers and nonlinear activation functions to learn linguistic and semantic patterns indicative of AI-generated fake news. The model's architecture is optimized through hyperparameter tuning, including adjustments to learning rate, batch size, and hidden layer dimensions. The choice of MLP ensures robustness and adaptability compared to traditional machine learning algorithms. This module establishes the foundation for accurate and scalable fake news classification.

5. Analyze and Prediction:

This module involves training the MLP model with preprocessed data and analyzing its predictive performance. During training, the model learns from both authentic and AI generated samples to understand the linguistic differences between them. Once trained, the system predicts whether a new article is genuine or fake based on its textual features. The analysis includes identifying key words, sentence structures, and tone patterns that influence the prediction. Visualization tools are used to interpret model results and understand prediction trends. This module ensures that the model can make accurate, data-driven decisions on unseen data.

6. Accuracy on Test Set:

In this module, the trained model is evaluated on the test dataset to measure its performance using standard metrics such as accuracy, precision, recall, and F1-score. The MLP classifier's results are compared against traditional models like Logistic Regression and Naïve Bayes to validate its superiority. Confusion matrices and ROC curves are used to assess the classifier's reliability and ability to distinguish between real and fake news. The evaluation helps determine how well the model generalizes to unseen text data. This module ensures that the system is robust and consistent before deployment.

7. Saving the Trained Model:

After achieving satisfactory accuracy, the trained MLP model is saved using serialization libraries such as pickle or joblib for future use. This enables quick loading of the model for real-time fake news detection without retraining. The saved model can be integrated into web applications, browser extensions, or automated fact-checking tools. Version control and proper documentation are maintained to ensure reproducibility. This module ensures that the detection system is deployment-ready, efficient, and scalable for realworld applications.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the model effectively distinguished AI-generated fake news from genuine news articles by leveraging both semantic text representations and external knowledge information. The integration of knowledge graphs significantly improved contextual understanding and fact verification capabilities.

The analysis showed that the graph-aware transformer architecture captured complex

relationships among entities, events, and contextual information, resulting in higher detection accuracy compared to conventional transformer-based fake news detection models.

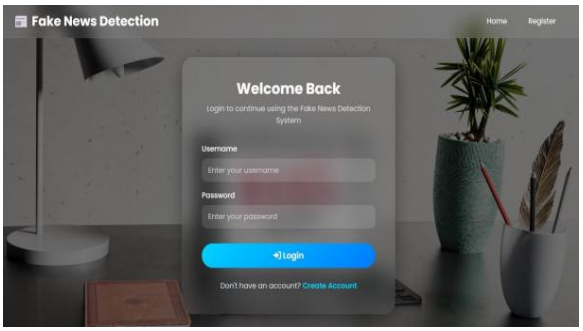


Fig. 2: Login



Fig. 3: Fake News Detector



Fig. 4: Detected Results



Fig. 5: Detected Results

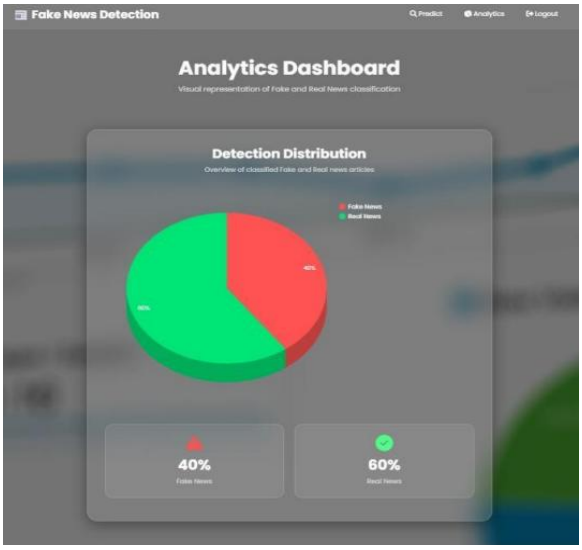


Fig. 6: Analytics Dashboard

V. CONCLUSION

The proposed study successfully demonstrates the effectiveness of an MLP-based classifier integrated with NLP techniques for detecting AI-generated fake news. By preprocessing textual data through tokenization, stop-word removal, and vectorization, the model accurately captures complex linguistic structures and semantic cues. The results indicate that the MLP classifier performs better than traditional machine learning approaches in identifying

fabricated news. The newly developed dataset, consisting of both human-written and AI-generated articles, provided a strong foundation for training and evaluation. The system achieved high accuracy and strong F1-scores, validating its reliability and robustness. Overall, this project highlights the potential of deep learning in preserving information integrity, promoting digital trust, and combating misinformation in the era of advanced AI content generation.

REFERENCES

[l] R. Zhao and T. Shi, "The impact of large language models on public security intelligence work and countermeasures research," J. Big Data Comput., vol. 2, no. 1, pp. 91—103, Mar. 2024.

[2] N. Bontridder and Y. Poullet, "The role of artificial intelligence in disinformation," Data Policy, vol. 3, p. e32, Jan. 2021.

[3] S. Kreps, R. M. McCain, and M. Brundage, "All the news that's fit to fabricate: Ai-generated text as a tool of media misinformation," J. Exp. Political sci., vol. 9, no. 1, pp. 104-117, 2022.

[4] M. R. Kondamudi, S. R. Sahoo, L. Chouhan, and N. Yadav, "A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches," J. King Saud Univ.-Comput. Inf.' Sci., vol. 35, no. 6, Jun. 2023, Art. no. 101571.

[5] A. Orhan, "Fake news detection on social media: The predictive role of university students' critical thinking dispositions and new media literacy," Smart Learn. Environments, vol. 10, no. l, p. 29, Apr. 2023.

[6] H. Schütze, C. D. Manning, and P. Raghavan, Introduction to Information Retrieval. Cambridge, U.K.: Cambridge Univ. Press, 2008.

AUTHORS Profile



Mr. ADALA AMARNATH has received his B.Tech in CSE from JNTUH, Hyderabad in 2013, M.Tech in Computer Science and Engineering from JNTUH, Hyderabad in 2016 and He has been dedicated to the teaching field for the last 6 years. At present he is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. ENUMULA BHAVYA SREE pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. AARADHYA SINGEWAR pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. ADEPU KALYAN pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. SHREYAN pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.