ISSN 2277-2685

IJESR/April. 2026/ Vol-16/Issue-2s1/43-52

Syed Ali Asrar *et. al.*, / International Journal of Engineering & Science Research

Full Length Article

KEGAT: A Knowledge-Enhanced Graph-Aware Transformer for Detecting AI-Generated Fake News

Syed Ali Asrar[1],Syed Akmal Uddin[2], Mohammed Ali Hussain[3], Sumrana Tabassum[4]

[1,2,3]BE Students; Department Of Computer Science Engineering ISL Engineering College Hyderabad India

[4]Assistant Professor ;Department Of Computer Science Engineering ISL Engineering College Hyderabad India

Mail Id;syedaliasrar19@gmail.com, syedakmal0808@gmail.com, mohammedalihussain1412@gmail.com, sumranaazmath10@gmail.com

Accepted 25-04-2026

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.

Keywords- *Fake News, MLP, NLP, GPT, Machine Learning.*

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 AI-generated 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.

Scope of the paper

The scope of this study is to develop an automated system for detecting AI-generated fake news using advanced machine learning techniques. The system

High Computational Complexity due to multiple integrated models (BERT, BiLSTM, TextCNN, Attention).

Requires Extensive Training Data and Computational Resources.

Difficult to Interpret because of hybrid deep learning architecture.

Risk of Overfitting when trained on small or imbalanced datasets.

applies Natural Language Processing (NLP) methods to preprocess textual data and extract meaningful features required for accurate classification. It is designed to identify complex linguistic patterns, writing styles, and contextual nuances commonly found in AI-generated content. A diverse dataset containing multiple news categories generated by GPT-4 is used to create a realistic environment for training and evaluating the model. The Multi-Layer Perceptron (MLP) classifier is selected for its capability to model nonlinear relationships and capture hidden patterns within the data. This system can be further extended for real-time monitoring of online news websites, blogs, and social media platforms. It also overcomes several limitations of traditional classifiers that fail to detect highly sophisticated synthetic text. Moreover, the project offers a scalable solution for fact-checking and news verification, while contributing to online information integrity and strengthening public trust in digital media platforms.

Existing System:

The existing system for detecting AI-generated fake news employs a hybrid deep learning approach known as the Global-Local News Detection Model. This architecture integrates BERT for contextual word embeddings, BiLSTM for capturing sequential dependencies, TextCNN for extracting local semantic patterns, and attention mechanisms for highlighting critical textual features. Together, these components enable the system to analyze both global and local linguistic cues within news articles, allowing it to identify subtle differences between authentic and fabricated content. By leveraging diverse feature extraction techniques, the model attempts to maximize accuracy in detecting AI-generated misinformation. The system operates by preprocessing news articles through standard NLP steps such as text cleaning, tokenization, and embedding generation. The preprocessed data is then passed through BERT to obtain deep contextual representations of words and sentences. BiLSTM layers capture long-term dependencies across sequences, while TextCNN layers extract fine-grained local features. The attention mechanism further refines this process by focusing on the most informative words or phrases. Finally, the outputs are combined and fed into a classification layer that predicts whether a news article is genuine or AI-generated. Although effective, the model is computationally expensive and complex, making it less suitable for large-scale or real-time applications.

Proposed System

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 GPT-4, 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.

Proposed System Advantages:

Simpler Architecture with Faster Training and Inference.
Efficient Handling of Text Features with NLP Preprocessing.
Lower Computational Cost compared to hybrid deep learning models.
Good Generalization with Proper Feature Engineering.
Scalable and Easily Deployable for Real-World Fake News Detection.

Literature Review

Title: FakeGPT: Fake News Generation, Explanation and Detection of Large Language Models
Author: Zhipeng Wu, Zhenyu Dai, Yanan Sun, et al.
Year: 2023
Description: 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.

Title: Event-Radar: Event-driven Multi-View Learning for Multimodal Fake News Detection
Author: Zihan Ma, Minnan Luo, Hao Guo, Zhi Zeng, Yiran Hao, Xiang Zhao
Year: 2024
Description: 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.

Title: A novel semantic deep learning approach to fake news detection
Author: J. Alghamdi, A. Author2 (et al.)
Year: 2024
Description: 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.

Title: A Practical Examination of AI-Generated Text Detectors for Large Language Models
Author: D. Ippolito, E. Wallace, K. Verstak, et al.
Year: 2024
Description: This empirical study systematically evaluates a range of AI-generated-text detectors in realistic (black-box) settings, testing watermarking, statistical detectors, and ML classifiers against un-watermarked LLM outputs. The findings reveal detection strengths and limits across LLM variants and adversarial generation strategies, and stress-test detector generalization — providing practical lessons on detector vulnerabilities and evaluation protocols that should inform how MLP-based fake-news detectors are trained and validated.

Title: Detecting AI-Generated Text: Factors Influencing Detectability with Current Methods (survey)
Author: K. C. Fraser, S. Author2, L. Author3 (survey authors)
Year: 2025
Description: This comprehensive 2025 survey reviews state-of-the-art AIGT (AI-generated text) detection techniques — watermarking, stylistic/statistical analysis, and machine learning approaches — and catalogs existing datasets, evaluation metrics, and adversarial challenges. It highlights arms-race dynamics (generation vs. detection), dataset creation pitfalls, and the need for robust cross-model evaluation. The survey is useful for guiding dataset construction (like your GPT-4 generated 42-category corpus), feature selection, and realistic evaluation setups for MLP classifiers.

Methodology

Modules Name:
Data Collection
Dataset
Data Preparation
Model Selection
Analyze and Prediction
Accuracy on test set
Saving the Trained Model

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.

Dataset:
The dataset consists of a mixture of real and AI-generated news articles to simulate real-world 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.

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.

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.

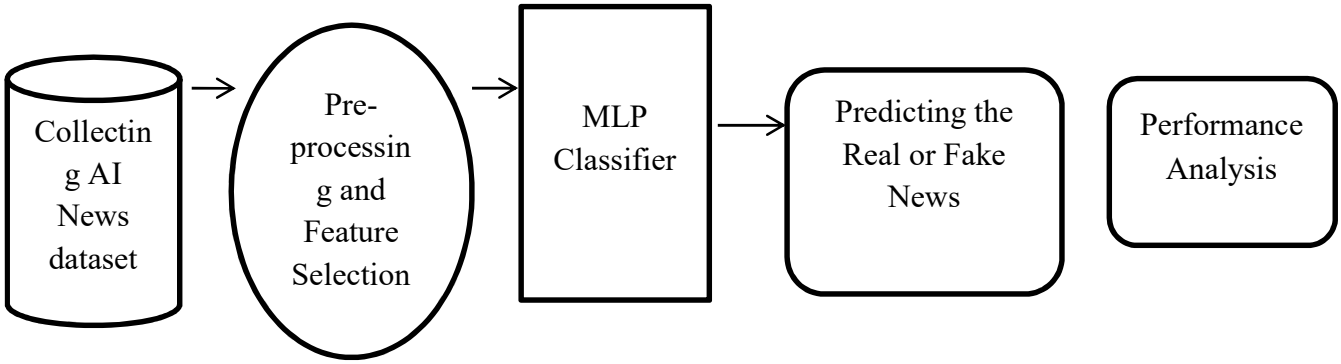
Analyze and Prediction:
This module involves training the MLP model with preprocessed data and analyzing its predictive performance. During training, the model learns from both authentic and AI-generated samples to understand the linguistic differences between them. Once trained, the system predicts 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.

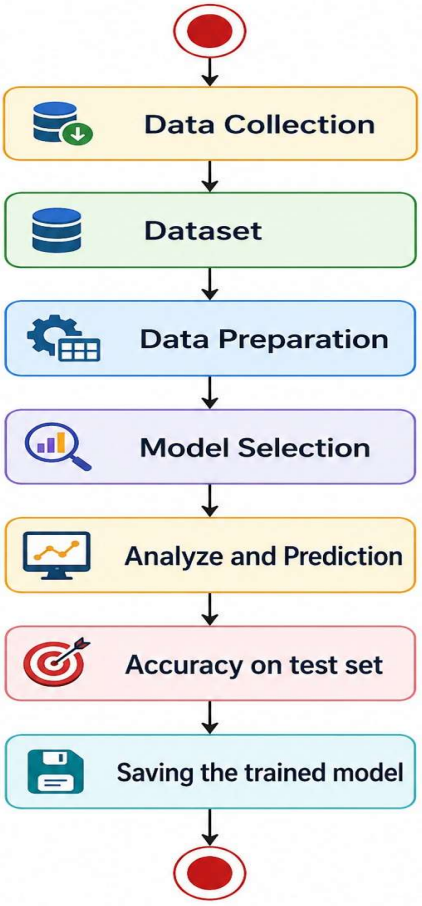
Accuracy on Test Set:
In this module, the trained model is evaluated on the test dataset to measure its performance using standard metrics such as accuracy, precision, recall, and F1-score. The MLP classifier's results are compared against traditional models like Logistic Regression and Naïve Bayes to 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

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 real-world applications..

Block Diagram:





.

Implementation

ALGORITHM USED

EXISTING TECHNIQUE: -

The Global-Local News Detection Model is a hybrid deep learning framework that integrates multiple architectures for robust text classification. BERT generates contextual embeddings that serve as the foundation for feature representation. These embeddings are passed into BiLSTM layers to capture

sequential dependencies and contextual flow within the text. In parallel, TextCNN extracts local n-gram features, identifying patterns within smaller text windows. An attention mechanism assigns greater weight to important tokens, improving interpretability and classification performance. Finally, the concatenated features are passed through dense layers to predict whether the input news article is authentic or AI-generated.

PROPOSED TECHNIQUE USED OR ALGORITHM USED:

The Multi-Layer Perceptron (MLP) classifier is a supervised feed forward neural network that maps input features to output categories through multiple layers of neurons. In this model, textual data is first processed using NLP techniques to convert raw news articles into numerical feature vectors. These vectors are fed into the MLP's input layer, followed by one or more hidden layers where nonlinear activation functions (e.g., ReLU, sigmoid) capture complex linguistic dependencies. The output layer generates predictions indicating whether the news is authentic or AI-generated. Training is performed using back-propagation and gradient descent to minimize classification error. This approach enables the MLP to learn subtle patterns in language, achieving reliable and accurate fake news detection.

Testing

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner.

Unit testing

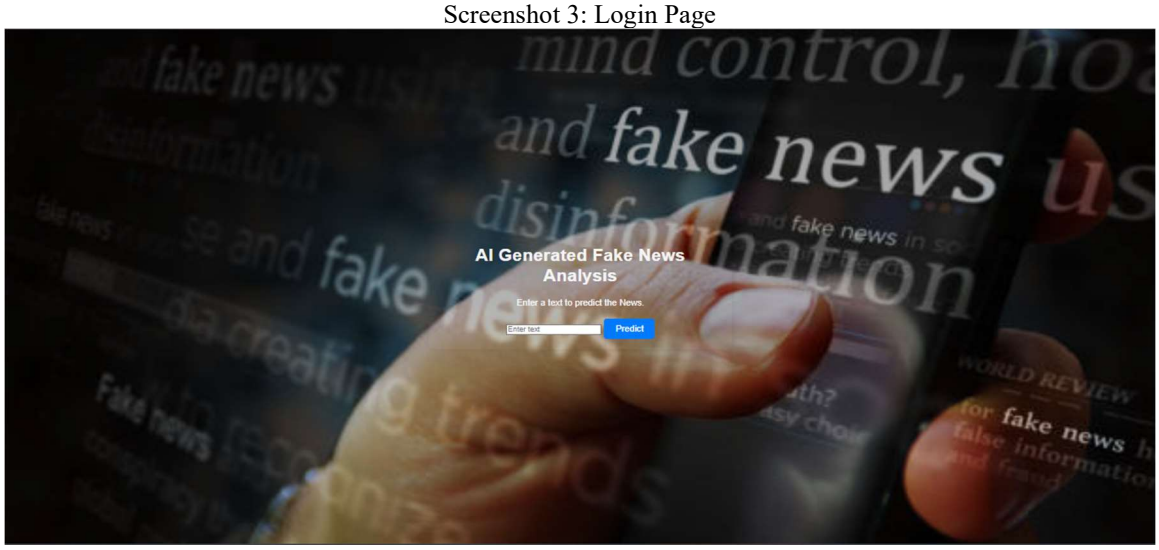
Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program input produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results

Results

Screenshot 1: Main Menu Page

Screenshot 2: Sign Up Page

Screenshot 3: Login Page



Screenshot 4: Data Input Page

Screenshot 5: Result 1 (FAKE)

Screenshot 6: Result 2 (REAL)

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.

Future Scope

In the future, the proposed fake news detection system can be enhanced by integrating advanced deep learning architectures such as Transformers and BERT for improved text understanding. Incorporating multimodal analysis that includes images, videos, and social media metadata can strengthen detection accuracy. The system can also be expanded to handle multilingual datasets, enabling fake news identification across different languages and regions. Real-time news stream analysis can be implemented to instantly flag AI-generated misinformation as it spreads online. Additionally, explainable AI (XAI) techniques can be introduced to make the prediction process more transparent and interpretable. Cloud-based deployment and API integration can make the system accessible for media agencies and the general public. Continuous model retraining with newly generated AI content will help maintain high accuracy as language models evolve. These enhancements would collectively make the system more adaptable, scalable, and efficient for future digital ecosystems.

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.

[7] L. Hu, S. Wei, Z. Zhao, and B. wu, "Deep learning forfakenews detection: A comprehensive survey," Al Open, vol. 3, pp. 133—155, Jan. 2022.

[8] H. Zhao, H. Chen, T. A. Ruggies, Y. Feng, D. Singh, and H.-J. Yoon, ¦Improving text classification with large language model-based data augmentation," Electronics, vol. 13, no. 13, p. 2535, Jun. 2024.

[91 K. Shu, A. Bhattacharjee, F. Alatawi, T. H. Nazer, K. Ding, M. Karami, and H. Liu, "Combating disinformation in a social media age," WIREs Data Mining Knowl. Discovery, vol. 10, no. 6, p. 1385, Nov. 2020.

[10] E. Shushkevich, M. Alexandrov, and J. Cardiff, "Improving multiclass classification of fake news using BERT-based models and ChatGPTaugmented data," Inventions, vol. 8, no. 5, p. 112, Sep. 2023.

[11] A. Naitali, M. Ridouani, F. Salahdine, and N. Kaabouch, "Deepfake attacks: Generation, detection, datasets, challenges, and research directions," Computers, vol. 12, no. 10, p. 216, Oct. 2023.

[12] P. Dhiman, A. Kaur, D. Gupta, S. Juneja, A. Nauman, and G. Muhammad, "GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection," Heliyon, vol. 10, no. 16, Aug. 2024, Art. no. e35865.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.

[14] A. Rakhlin, "Convolutional neural networks for sentence classification," GitHub, vol. 6, p. 25, Jan. 2014.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., vol. 30, Jun. 2017, pp. 5998—6008.

[16] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, no. 3, pp. 379-423, Jul. 1948.

[17] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Proc. Adv. Neural Inf Process. syst., vol. 26, Dec. 2013, pp. 3111-3119.

[18] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in Proc. Conf Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1532-1543.

[19] J. Elman, "Finding structure in time," Cognit. Sci., vol. 14, no. 2, pp. 179-211, Jun. 1990.

[20] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 157-166, Mar. 1994.

[21] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, vol. 1, Minneapolis, MN, USA, Jan. 2018, pp. 4171-4186

[22] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI, San Francisco, CA, USA, Tech. Rep., 2018.

[23] T. B. Brown et al., "Language models are few-shot learners," in Proc. Adv. Neural In/ Process. syst., Jan. 2020, pp. 1877-1901.

[24] Y. Zhou, T. Shen, X. Geng, G. Long, and D. Jiang, "ClarET: Pretraining a correlation-aware context-to-event transformer for event-centric generation and classification," in Proc. 60th Annu. Meeting Assoc. Comput. Linguistics, 2022, pp. 2559-2575.

[25] Y. Zhou, X. Geng, T. Shen, G. Long, and D. Jiang, "EventBERT: A pretrained model for event correlation reasoning," in Proc. ACM Web Conf. , Apr. 2022, pp. 850-859

[26] Y. Zhou, X. Geng, T. Shen, J. Pei, W. Zhang, and D. Jiang, "Modeling event-pair relations in external knowledge graphs for script reasoning," in

Proc. Findings Assoc. Comput. Linguistics, ACL-IJCNLP, Jan. 2021, pp. 4586-4596.

[27] S. Dudhmande, S. Golliwar, A. Bhagwat, R. Ghiya, and A. Bhade, "Textual compression using lamini-LM," Int. Res. J. Adv. Eng. Manage. (IRJAEM), vol. 2, no. 5, pp. 1536-1540, May 2024.

[28] H. Lei, X. Liu, G. Niu, Y. Zhou, and Y. Zhou, "Generative Al authorship verification based on ChatGLM," in Proc. Work. Notes CLEF, 2024, PP. 1-6.

[29] A. Zeng et al., "ChatGLM: A family of large language models from GLM130B to GLM-4 all tools," 2024, arXiv.•2406.12793.

[30] J.-B. Alayrac et al., "Flamingo: A visual language model for few-shot learning," in Proc. Conj: Neural Inf. Process. Syst., Jan. 2022, pp. 1—21.

[31] Z.-J. Yang, T.-F. Feng, and X.-G. Wu, "On the possibility of mixed axion/neutralino dark matter in specific SUSY DFSZ axion models," 2023, arXiv.•2303.11645.

[32]] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in Proc. 20th Int. Conf. World Wide Web, Mar. 2011, pp. 675—684. [331 R. Labaca-Castro, "Generative adversarial nets," in Proc. Adv. Neural Inf.' Process. syst., vol. 27, Jan. 2023, pp. 73-76.

[33] X. Li and Y. Zhou, "Disentangled and robust representation learning for bragging classification in social media," in Proc. IEEE Int. Conj: Acoust., Speech Signal Process. (ICASSP), Jun. 2023, pp. 1-5.

[34] K. M. DSouza and A. M. French, "Fake news detection using machine learning: An adversarial collaboration approach," Internet Res., vol. 34, no. 5, pp. 1664—1678, Sep. 2024.

[35] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," Int. J. Inf Manage. Data Insights, vol. l, no. l, Apr. 2021, Art. no. 100007.

[36] J. Alghamdi, Y. Lin, and S. Luo, "The power of context: A novel hybrid context-aware fake news detection approach," Information, vol. 15, no. 3, p. 122, Feb. 2024.

[37] R. Shao, T. Wu, and Z. Liu, "Detecting and grounding multi-modal media manipulation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 6904-6913.

[38] F. Shan, H. Sun, and M. Wang, "Multimodal social media fake news detection based on similarity inference and adversarial networks," Comput., Mater: Continua, vol. 79, no. 1, pp. 581—605, 2024.

[39] S. Wachter, B. Mittelstadt, and C. Russell, "Do large language models have a legal duty to tell the truth?" SSRN Electron. J., vol. 11, no. 8, 2024, Art. no. 240197.

[40] R. Misra. (2022). News Category Dataset. [Online]. Available: https:// huggingface.co/datasets/heegyu/news-category-dataset