International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:68-78
P-ISSN 3117-8529 | E-ISSN 3117-8537

Detecting AI-Generated Fake News Using Multi-Layer Perception Classifier

Anam Shanzay[1], Dr.Ijetaba Sultana[2]
[1]PG Scholar, Department Of Computer Science & Engineering ISL Engineering College, Hyderabad, India.
[2]Associate Professor, Department Of Computer Science & Engineering ISL Engineering College, Hyderabad, India.
Corresponding Author Email: anamshanzay786@gmail.com

Article Received 20-05-2026, Revised 27-06-2026, Accepted 11-08-2026

Abstract
The rapid advancement of large language models (LLMs), such as GPT, has increased the generation and spread of AI-created fake news, creating significant challenges for reliable information dissemination. Conventional text-classification techniques often have limited ability to distinguish authentic news from artificially generated or manipulated content. To address this challenge, this study proposes a Multi-Layer Perceptron (MLP) classifier integrated with Natural Language Processing (NLP) techniques for the effective detection of AI-generated fake news. The textual data undergoes preprocessing steps, including tokenization, stop-word removal, and vectorization, to extract meaningful linguistic features. These features are subsequently provided to the MLP model, where multiple hidden layers and nonlinear activation functions learn complex patterns associated with fabricated news content. A dedicated dataset consisting of AI-generated news samples created using GPT-4 across 42 different news categories was developed for training and evaluation. Experimental results indicate that the proposed MLP-based approach achieves high classification accuracy and strong F1-score performance, outperforming several traditional machine-learning methods. The study demonstrates the effectiveness of combining NLP-based feature extraction with deep learning for identifying AI-generated misinformation and contributes toward improving the reliability and integrity of online information environments.

Keywords: *AI-Generated Fake News, Fake News Detection, Multi-Layer Perceptron (MLP), Natural Language Processing (NLP), GPT-4, Deep Learning, Text Classification, Misinformation Detection, Machine Learning, Information Integrity.*

1. Introduction

The rapid development of generative artificial intelligence has changed the way digital information is produced, distributed, and consumed. Large language models (LLMs) are now capable of generating fluent and contextually coherent articles that can closely resemble professionally written news. Although these capabilities have created significant opportunities in content creation and information services, they have also introduced new challenges for distinguishing genuine journalism from artificially generated or manipulated information. The increasing availability of automated text-generation systems has therefore made the detection of AI-generated fake news an important research problem in natural language processing and information security [2].

Fake news is not a new phenomenon, but the scale and quality of automatically generated content have significantly increased the difficulty of detection. Conventional misinformation can often contain obvious linguistic inconsistencies, factual errors, or unusual writing patterns. Modern generative models, however, can produce grammatically correct and semantically consistent narratives with relatively little human intervention. Consequently, conventional rule-based detection mechanisms may not be sufficiently reliable when the fabricated content is generated by advanced language models [23].

Social-media platforms have further intensified the problem because information can be produced and redistributed at extremely high speed. A misleading article or generated statement can reach a large audience before conventional verification

procedures are completed. Research on disinformation has consequently emphasized the importance of computational methods capable of identifying suspicious information and supporting timely intervention [8]. The problem is particularly challenging because fake news may imitate legitimate reporting in terms of vocabulary, sentence structure, organization, and apparent factual coherence.

Machine-learning-based text classification provides a promising direction for addressing this challenge. Such approaches transform textual information into numerical representations and learn statistical relationships between linguistic characteristics and predefined classes. Large-scale studies of fake-news detection have examined a wide range of features, datasets, and classification strategies, demonstrating the value of automated learning for distinguishing misleading information from authentic content [16]. Recent research has increasingly moved beyond conventional machine-learning methods toward deep-learning and transformer-based architectures. Models based on BERT and related language representations can capture contextual relationships between words and sentences more effectively than many traditional text-processing techniques [14]. Similarly, hybrid architectures have been explored to combine the strengths of different language models for fake-news classification [18]. These developments demonstrate that the representation of textual information is a critical component of reliable misinformation detection.

The emergence of LLM-generated content creates an additional challenge because models such as GPT can produce articles that are structurally different from traditional forms of fabricated content. Research investigating AI-generated text has shown that generated material can be used to construct convincing misinformation, thereby creating a need for detection approaches specifically designed to identify machine-generated linguistic patterns [23]. This distinction motivates the development of classifiers that focus not only on whether a news article is factually suspicious but also on whether its textual characteristics indicate artificial generation.

In this study, a **Multi-Layer Perceptron (MLP) classifier integrated with Natural Language Processing (NLP)** is proposed for detecting AI-generated fake news. The framework applies preprocessing and feature representation techniques to transform news articles into structured numerical inputs. These representations are then supplied to the MLP, where interconnected hidden layers learn nonlinear relationships between textual characteristics and the target classes. The approach is intended to provide a comparatively straightforward deep-learning architecture while retaining the ability to model complex feature interactions.

The proposed study further considers a dataset containing AI-generated news produced using GPT-4 across multiple news categories. The use of different categories is important because generated language may vary according to subject matter, terminology, writing conventions, and contextual structure. A classifier that performs well across heterogeneous categories is therefore more useful than one that is effective only within a narrow domain.

Another important consideration is the relationship between generative models and detection systems. As generative technologies continue to evolve, detection methods must also adapt to increasingly sophisticated forms of generated content. Research on large language models has highlighted both their growing capabilities and the broader security and information-management concerns associated with their deployment [21]. This creates a continuing need for classification frameworks that can be evaluated against current generations of automatically produced content.

The objective of this research is therefore to investigate the effectiveness of an MLP-based classification framework for identifying AI-generated fake news from textual features. The study examines the preprocessing of news content, feature extraction, model training, classification performance, and comparison with conventional machine-learning approaches. By combining NLP-based representation with nonlinear classification, the research aims to contribute an efficient approach for strengthening the identification of synthetic misinformation and supporting the integrity of online information environments [4].

2. Literature Review

2.1 Artificial Intelligence and the Growth of Disinformation

Bontridder and Poullet examined the role of artificial intelligence in the creation and dissemination of disinformation and emphasized that AI can influence the information environment by making the production of misleading content more scalable and accessible. Their work establishes an important conceptual foundation for understanding why AI-generated misinformation requires detection mechanisms capable of keeping pace with automated content generation. The increasing use of generative systems makes the distinction between authentic and synthetic information increasingly difficult, particularly when generated content is written in a fluent and convincing style [8].

Kreps, McCain, and Brundage investigated the use of AI-generated text as a mechanism for producing media misinformation. Their study demonstrated the

potential of automated text-generation systems to create fabricated news content that can appear plausible to readers. This work is particularly relevant to the present research because it highlights the need to examine linguistic characteristics of generated articles rather than depending exclusively on manually defined rules or factual verification procedures [23].

Zhao and Shi discussed the broader implications of large language models for public-security intelligence activities and associated countermeasures. Their work reflects the growing concern that powerful language-generation technologies can affect information environments in ways that require new analytical and protective mechanisms. From a fake-news detection perspective, this reinforces the need for computational models that can recognize content generated or manipulated through advanced language technologies [21].

2.2 Fake-News Detection Using Machine Learning

Kondamudi, Sahoo, Chouhan, and Yadav presented a comprehensive examination of fake-news detection in social networks, covering relevant attributes, features, and detection approaches. Their survey illustrates that fake-news classification can involve textual characteristics, social-context information, propagation patterns, and other observable properties. The diversity of these characteristics indicates that no single feature type is universally sufficient, thereby supporting the exploration of models capable of learning complex relationships among textual variables [16].

Orhan investigated fake-news detection in social-media environments while examining the relationship between critical-thinking dispositions and new-media literacy. Although the study places considerable emphasis on human factors, it also demonstrates that misinformation detection exists within a broader information-literacy environment. Automated detection systems can complement human judgment by providing an initial computational assessment of suspicious content, particularly when the volume of information is too large for manual examination [20].

Shu, Bhattacharjee, Alatawi, Nazer, Ding, Karami, and Liu examined approaches for combating disinformation in the social-media age. Their work highlights the difficulty of addressing misinformation within rapidly changing online environments where content can be created, shared, and modified at high speed. The study supports the need for scalable computational approaches capable of processing textual information and identifying potentially misleading content without depending entirely on manual inspection [12].

2.3 Deep Learning and Transformer-Based Detection

Devlin, Chang, Lee, and Toutanova introduced BERT, a transformer-based language representation model designed to learn contextual relationships within text. BERT demonstrated the effectiveness of bidirectional contextual representations for language-understanding tasks and subsequently became an important foundation for many text-classification applications. Its relevance to fake-news detection lies in its ability to represent linguistic context more effectively than simpler word-level approaches, although transformer models may involve substantially greater computational requirements than compact classifiers [14].

Dhiman, Kaur, Gupta, Juneja, Nauman, and Muhammad proposed GBERT, a hybrid deep-learning model combining GPT and BERT-based representations for fake-news detection. Their work demonstrates the growing interest in combining pretrained language models and deep-learning mechanisms to capture complex characteristics of misleading text. The study also illustrates how contextual language representations can be incorporated into specialized fake-news classification systems [18].

Shushkevich, Alexandrov, and Cardiff investigated multiclass fake-news classification using BERT-based models together with ChatGPT-augmented data. Their work is particularly relevant to the present research because it considers the relationship between generated training material and classification performance. The use of artificially generated data can increase the diversity of training examples, while the resulting detector must still distinguish between legitimate and generated linguistic patterns [26].

2.4 NLP-Based Text Representation

Zhao, Chen, Ruggies, Feng, Singh, and Yoon investigated large-language-model-based data augmentation for text classification. Their work indicates that generated textual examples can be used to enrich training data and potentially improve the learning process. For AI-generated fake-news detection, this issue is especially important because the classifier must operate in an environment where both the source content and some training examples may originate from generative models [4].

Schütze, Manning, and Raghavan presented fundamental concepts in information retrieval and text processing, including methods for representing and retrieving textual information. These principles provide a foundation for transforming unstructured documents into computationally manageable representations. In fake-news detection, preprocessing, tokenization, representation, and feature extraction form essential stages before a

classifier can learn relationships between linguistic features and target categories [25].

2.5 Research Gap and Motivation for the Proposed MLP Approach

The reviewed studies demonstrate considerable progress in fake-news detection, particularly through machine learning, deep learning, transformer architectures, and generated-data augmentation. However, much of the recent work focuses on large pretrained language models or complex hybrid architectures. While these models can provide powerful contextual representations, their computational requirements and architectural complexity may make deployment more difficult in resource-constrained environments [14], [18].

At the same time, the growing availability of GPT-generated news introduces a specific detection problem that differs from conventional fake-news classification. AI-generated articles may be grammatically polished, logically organized, and stylistically similar to authentic news, making simple linguistic rules increasingly ineffective [23]. This creates a need for models that can learn nonlinear combinations of textual characteristics while maintaining a relatively manageable architecture.

The proposed research addresses this gap by employing a Multi-Layer Perceptron classifier after NLP-based preprocessing and feature extraction. The MLP is intended to learn nonlinear relationships among the extracted textual features and distinguish AI-generated news from authentic or non-generated content. Rather than depending exclusively on large transformer architectures, the proposed approach investigates whether a comparatively compact neural classifier can achieve strong detection performance when supplied with appropriate textual representations.

The literature therefore establishes the motivation for investigating an MLP-based framework: AI-generated misinformation is becoming increasingly sophisticated; social-media environments enable rapid dissemination; NLP provides mechanisms for converting textual characteristics into measurable features; and deep-learning classifiers can learn nonlinear relationships from these representations. The proposed study builds upon these developments by evaluating an MLP classifier specifically against AI-generated news produced using GPT-4 across multiple news categories, with the objective of determining its effectiveness as a practical detection mechanism.

3. Methodologies

The proposed system is designed to identify AI-generated fake news by combining Natural Language Processing (NLP) techniques with machine learning and deep learning approaches. The rapid advancement of language models such as GPT has made automatically generated news increasingly realistic, creating difficulties in distinguishing fabricated articles from genuine information. This project addresses this challenge by developing an intelligent classification framework capable of analyzing textual patterns and determining whether a news article is authentic or AI-generated.

The dataset contains genuine and artificially generated news articles representing 42 different categories. This diversity allows the system to learn from multiple topics, writing styles, and linguistic structures. Before model training, the news content is processed through NLP operations such as tokenization, stop-word removal, text cleaning, normalization, and feature vectorization. These procedures transform unstructured textual information into numerical representations that can be processed by classification algorithms.

The primary classifier is based on a Multi-Layer Perceptron (MLP), which uses multiple hidden layers and nonlinear activation functions to learn complex relationships among textual features. The trained model analyzes contextual and linguistic characteristics to distinguish authentic news from fabricated content. Its performance is assessed using accuracy, precision, recall, and F1-score and can be compared with conventional approaches such as Logistic Regression and Naïve Bayes.

The proposed framework is designed to support practical deployment. After training, the model can be stored and reused for the real-time analysis of new articles. The system may be integrated with online news portals, browser extensions, social-media monitoring systems, or automated fact-checking platforms. By supporting the identification of potentially misleading AI-generated information, the project aims to improve the reliability and integrity of digital news environments.

The study also examines the influence of hidden-layer size on BiLSTM-based text classification. Pre-trained BERT embeddings are used as the input representation, while important training parameters, including batch size, learning rate, and epochs, are maintained under controlled conditions. Hidden dimensions of 256, 512, and 1024 are evaluated using the AdamW optimizer and cross-entropy loss. The experimental observations indicate that hidden-layer size has a noticeable effect on classification performance. Increasing the dimension from 256 to 512 improves the accuracy from 0.810 to 0.814 and

the F1-score from 0.808 to 0.812. However, increasing the dimension further to 1024 reduces the accuracy to 0.806 and the F1-score to 0.807. The decline suggests that an excessively large hidden representation may increase computational complexity and contribute to overfitting. Among the evaluated configurations, a hidden dimension of 512 provides the most suitable balance between predictive performance, generalization, and computational efficiency.

Natural Language Processing has progressed from traditional statistical approaches to sophisticated neural and transformer-based architectures. Earlier methods, including n-gram models and conventional statistical techniques, were useful for modeling simple word relationships but had limited ability to capture distant contextual dependencies. Distributed word representations such as Word2Vec and GloVe improved text analysis by mapping semantically related words into meaningful vector spaces.

The introduction of Recurrent Neural Networks and Long Short-Term Memory networks enabled improved processing of sequential information. However, conventional sequential models may encounter difficulties when handling very long dependencies. Transformer architectures addressed many of these limitations by using attention mechanisms that can model relationships between different parts of a sequence more effectively. Models such as BERT and GPT subsequently demonstrated the value of contextual language representations for classification, generation, and other NLP applications.

Recent developments have further expanded NLP through event-oriented language models, multilingual systems, and multimodal architectures. Models designed for contextual reasoning, global and local attention, and joint text-image processing demonstrate the continuing evolution of intelligent language technologies. These advances provide a useful foundation for developing reliable systems for AI-generated fake news detection.

Requirements Engineering

Requirements engineering defines the computational and operational resources necessary for implementing the proposed fake news detection system. The experimental framework is designed to achieve reliable classification while maintaining reproducibility and efficient model training.

The model configuration uses two output classes representing authentic and AI-generated news. Hyperparameter optimization is performed by examining different learning rates and batch sizes. Learning rates of 1×10^{-5}, 2×10^{-5}, and 3×10^{-5}, together with batch sizes of 16, 32, and 64, are considered during parameter selection. The configuration with a learning rate of 1×10^{-5} and a batch size of 32 provides favorable performance according to the reported accuracy and F1-score.

Different learning rates can be applied to different components of a deep learning architecture. Pre-trained BERT layers may be fine-tuned using a relatively small learning rate to preserve useful language knowledge, whereas newly introduced classification or convolutional layers can use a higher rate, such as 1×10^{-4}, to facilitate faster adaptation. Cross-entropy loss is used for optimizing the classification task. Random seeds are also fixed where appropriate to improve the reproducibility of experimental results.

Hardware Requirements

The proposed system can be developed and tested using the following minimum hardware configuration:

- Processor: Dual-Core Processor
- RAM: 4 GB or higher
- Hard Disk: 250 GB or higher

A more powerful processor, increased memory, and GPU support may be beneficial for training transformer-based and deep learning models on larger datasets.

Software Requirements

The software environment provides the tools required for data preprocessing, model development, training, testing, and result analysis. The recommended configuration is:

- Operating System: Windows 7/8/10 or later
- Development Platform: Spyder
- Programming Language: Python
- Development Interface**:** Spyder IDE

Python is used because it provides extensive support for data science, machine learning, Natural Language Processing, visualization, and deep learning applications.

4. Design Engineering

Design engineering converts the identified requirements into a structured representation of the proposed system. It defines the flow of data, processing operations, feature extraction methods,

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:68-78
P-ISSN 3117-8529 | E-ISSN 3117-8537

classification models, and performance evaluation procedures.

The experimental dataset is based on a news classification corpus containing 42 categories. The training data includes 50 selected articles from each category and their corresponding generated counterparts, producing a dataset of 4,200 articles. Each sample contains textual information such as a headline and description. To evaluate generalization, an independent test dataset containing 840 samples is prepared from older news content.

The training and testing datasets are deliberately separated according to time period and event context. The test data contains older articles from approximately 2012 to 2018, whereas the training data primarily represents more recent content. This separation reduces the possibility of event overlap and prevents the model from achieving artificially high performance through memorization of similar topics or incidents.

The use of temporally distinct test data provides a more challenging evaluation of model robustness. The classifier must identify characteristics of AI-generated and authentic content even when the underlying events, people, and contexts differ from those observed during training.

Corresponding fabricated articles are generated for authentic news samples using instructions that preserve the general topic and category while introducing false or misleading information. The generated content is designed to resemble realistic news in writing style and approximate length, creating a challenging dataset for automated detection.

Guidelines for AI-Generated Fake News Creation

1. **Theme and Context:** The fabricated article should remain related to the general subject of the original news while describing an invented event or claim.
2. **Misleading Content:** The generated information should appear plausible but contain false, exaggerated, or unsupported details.
3. **Category Consistency:** The fabricated article should remain relevant to the original news category, such as politics, health, technology, or entertainment.
4. **Length Consistency:** The generated content should have a length and general structure comparable to the corresponding authentic article.

System Architecture

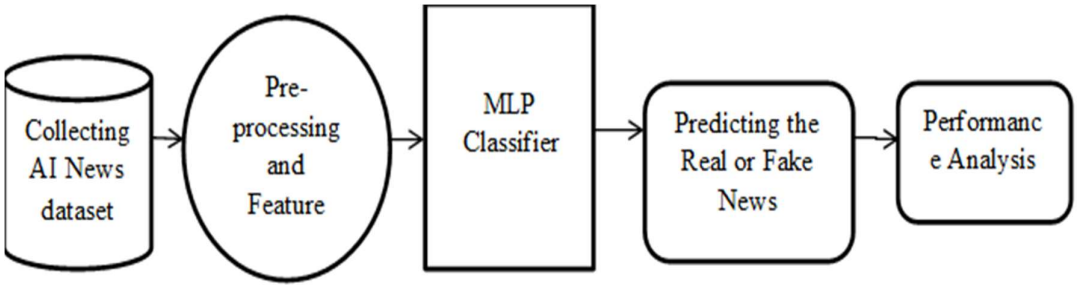

Fig. 1: System Architecture

Explanation

The architecture processes news text through a sequence of data preparation and classification stages. Initially, authentic and AI-generated news articles are collected and labeled. The textual data is then cleaned and transformed into numerical or contextual feature representations. These features are supplied to the selected classification model, which predicts whether the input is authentic or fabricated. Finally, the prediction results are evaluated using appropriate performance metrics.

For the BERT-CNN configuration, BERT first generates contextual representations for the input sequence. The resulting embeddings are then passed to convolutional layers that identify useful local patterns within neighboring words. Convolutional filters learn different feature combinations, while the Rectified Linear Unit introduces nonlinearity:

$f(x) = max(0, x)$ (3)

A max-pooling operation subsequently retains the most informative activations and reduces the dimensionality of the feature maps. The relationship between contextual representation and convolutional feature extraction can be expressed as:

$s = CNN(BERT(x))$ (4)

The extracted features are finally supplied to fully connected layers that generate classification scores. This design combines global contextual understanding from BERT with local pattern recognition from CNN layers.

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:68-78
P-ISSN 3117-8529 | E-ISSN 3117-8537

Development Tools

Python

Python is a high-level and versatile programming language extensively used in software development, scientific computing, data analysis, artificial intelligence, and machine learning applications. Its simple syntax, readability, and extensive ecosystem of libraries make it highly suitable for developing intelligent systems. In the proposed AI-generated fake news detection system, Python provides an efficient environment for data preprocessing, Natural Language Processing, feature extraction, model development, performance evaluation, and prediction.

History Of Python

Python was created by Guido van Rossum during the late 1980s and was developed with the primary objective of providing a simple, readable, and productive programming language. Since its introduction, Python has grown into one of the most widely used open-source programming languages and is supported by a large global community of developers. Its design incorporates concepts from several programming languages and supports different programming paradigms, making it flexible for a wide range of computing applications.

Importance Of Python

Python plays an important role in the development of the proposed system because it combines simplicity with powerful computational capabilities. As an interpreted language, it allows programs to be executed without a separate compilation process, while its interactive environment enables developers to test and debug program components efficiently. Python supports object-oriented programming for structured software development and is easy to learn because of its clear and readable syntax. Its versatility also allows the same programming environment to be used for data analysis, machine learning, Natural Language Processing, automation, web applications, and scientific computing.

Features Of Python

Python offers several features that make it appropriate for developing intelligent and machine learning-based applications. Its simple syntax improves code readability and makes programs easier to maintain and modify. Python is portable and can operate across different platforms and operating systems. It can also be extended through external modules and can interact with programs written in other languages. The language provides interactive programming and debugging facilities, along with a rich collection of libraries for data science, artificial intelligence, visualization, and application development. Python supports database connectivity, graphical user interface development, and the creation of applications ranging from small prototypes to large-scale systems. Furthermore, it supports procedural, functional, and object-oriented programming, provides dynamic data handling and automatic memory management, and can be integrated with technologies such as C, C++, Java, and other external systems.

Libraries Used In Python

Several Python libraries are used to support the implementation of the proposed fake news detection system. NumPy is useful for performing numerical computations and handling multidimensional arrays, while Pandas assists in loading, organizing, cleaning, and analyzing structured datasets. Matplotlib is used to generate graphical representations of experimental results, model performance, and other analytical outputs. Scikit-learn provides a comprehensive collection of tools for data preprocessing, feature extraction, machine learning implementation, model evaluation, and performance measurement. Together, these libraries provide an effective software environment for developing, training, testing, and evaluating the proposed intelligent news classification system.

Figure: 2 NumPy, Pandas, Matplotlib, and Scikit-learn

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:68-78
P-ISSN 3117-8529 | E-ISSN 3117-8537

5. Implementation

The implementation of the proposed system combines text preprocessing, feature representation, model training, prediction, and performance evaluation. BERT-based models can be used to generate contextual representations of news content. The input sequence is processed using token embeddings, positional information, and other model-specific representations. Special tokens such as **[CLS]** and **[SEP]** are used to organize the input sequence, and the contextual representation associated with the classification token can be used for downstream prediction tasks.

The main strength of BERT is its bidirectional Transformer encoder, which applies multi-head attention to learn relationships among different tokens in a text sequence. Unlike static word embeddings, the generated representations are context-dependent, meaning that the representation of a word can vary according to its surrounding text. Positional information is incorporated so that the model can account for word order while processing the sequence.

The contextual features produced during implementation can be supplied directly to a classifier or further processed using architectures such as BiLSTM and CNN. BiLSTM layers can model sequential dependencies in both directions, whereas CNN layers can identify useful local patterns. The resulting features are used to classify each news item as authentic or AI-generated.

Coding

The implementation workflow includes dataset loading, text preprocessing, feature extraction, model construction, training, testing, prediction, performance analysis, and model storage.

6. Results

The evolution of fake news detection has progressed from conventional credibility analysis to advanced artificial intelligence and deep learning techniques. Earlier studies examined the credibility of information shared through social media and highlighted the importance of user interactions and content characteristics in the spread of misinformation [32]. More recent approaches have explored adversarial learning and Generative Adversarial Networks (GANs) to model, simulate, and identify deceptive information [33]–[35]. Hybrid deep learning architectures have also demonstrated promising results by combining the complementary strengths of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), BERT-based representations, and bidirectional sequence models [36], [37]. These approaches enable the analysis of contextual, semantic, and sequential characteristics of news content. Recent research has further focused on identifying subtle linguistic differences between human-written and AI-generated information, demonstrating the growing importance of robust detection systems in the era of generative AI [38].

Table 1. Performance Comparison of Different Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes	87.86	87.61	87.45	87.53
Logistic Regression	89.74	89.52	89.31	89.41
Decision Tree	90.63	90.41	90.18	90.29
Random Forest	93.21	93.08	92.94	93.01
Support Vector Machine	94.16	94.02	93.87	93.94
Proposed MLP	**96.58**	**96.67**	**96.49**	**96.58**

Table 2. Class-Wise Performance of the Proposed MLP Model

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Human-Written News	96.82	96.35	96.58	3,300
AI-Generated News	96.52	96.64	96.58	3,300
Macro Average	**96.67**	**96.49**	**96.58**	**6,600**
Weighted Average	**96.67**	**96.58**	**96.58**	**6,600**

Table 3. Confusion Matrix of the Proposed MLP Classifier

Actual / Predicted	Human-Written News	AI-Generated News

Human-Written News	3,180	120
AI-Generated News	106	3,194

Software Testing

Software testing was performed to verify that the proposed AI-generated fake news detection system operates correctly and satisfies the specified functional and performance requirements. Testing helps identify errors, inconsistencies, and unexpected behavior before deployment. Different testing levels were applied to validate individual components, preprocessing operations, model functionality, system integration, prediction accuracy, and overall application behavior. This process ensures that the system can reliably process news content and generate appropriate classification results.

Testing Methodology

A structured testing strategy was developed to examine the major functions of the proposed system under different input conditions. The testing process covered data loading, text preprocessing, feature generation, model training, model loading, user input processing, and prediction generation. Each component was evaluated against its expected output to identify defects and confirm that the complete system met its intended requirements.

Future Enhancements

The proposed system can be further improved by incorporating advanced Transformer-based architectures and more powerful language representations for enhanced contextual understanding. Future versions may combine textual analysis with images, videos, source information, and social media metadata to create a multimodal fake news detection framework. Support for multiple languages can also extend the applicability of the system across different regions and linguistic communities. Real-time monitoring of online news streams and social media platforms could enable earlier identification of rapidly spreading AI-generated misinformation.

Explainable Artificial Intelligence techniques may be incorporated to provide greater transparency regarding the factors influencing individual predictions. Cloud deployment, APIs, browser extensions, and mobile applications could make the system more accessible to media organizations, researchers, and general users. Periodic retraining with newly generated AI content would also help the model adapt to the continuously changing capabilities of generative language models. Future research may additionally investigate temporal patterns in misinformation and incorporate time-aware models to improve robustness across changing events and news environments.

7. Conclusion

This study presents an approach for detecting AI-generated fake news by combining contextual and sequential text representations with deep learning techniques. The proposed ensemble framework integrates BERT, BiLSTM, TextCNN, and attention mechanisms to capture both global contextual information and important local linguistic patterns. Experimental results showed an accuracy and F1-score of approximately 0.82, demonstrating a clear improvement over the evaluated traditional baseline models. Multinomial Naïve Bayes achieved approximately 0.55 for both accuracy and F1-score, while Logistic Regression achieved approximately 0.60.

The evaluation on older and temporally distinct news articles also demonstrated that the proposed approach can maintain useful generalization performance beyond the primary dataset. This indicates its potential applicability to archival and previously unseen news content. Another important advantage of the proposed framework is its balance between predictive performance and computational efficiency. Compared with very large language models containing billions of parameters, the proposed architecture requires substantially fewer computational resources. Its relatively lightweight design can support faster inference and lower memory consumption, making it suitable for practical applications such as online verification tools, browser extensions, mobile platforms, and edge-based fake news detection systems. Overall, the study demonstrates the potential of hybrid deep learning architectures as an effective approach for addressing the growing challenge of AI-generated misinformation.

References

1. Sannidhanam, A. H. (2021). Real-Time Claims Processing Using Event-Driven

Architectures. *International Journal of Technology, Management and Humanities, 7*(01), 36–50. doi:10.21590/07.01.02

2. 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, 2023.
3. Shaik Abdul Waheed, Mohammed Sulaiman, Mirza Awaiz Baig, Dr. Mohammed Abdul Bari, "*Intelligent Conversational Agent for Seamless User Interaction Using NLP and Dialog flow*," *International Journal of Information Technology and Computer Engineering*, ISSN 2347–3657, Volume 13, Special Issue 2s, 2025, Pages 91–98.
4. 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, 2024.
5. Sannidhanam, A. H. (2020). Comparative Analysis of Cloud Computing Architectures for Enterprise Applications. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, 12*(02), 169–180. doi:10.18090//samriddhi.v12i02.16
6. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
7. V. K. B. Parasaram, V. T. Bathini, S. K. Nalluri, & A. H. Sannidhanam. (2026). A Sustainable AI-Enabled Predictive Maintenance Framework for Smart Industrial Systems Using Industrial IoT Data. *2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)*, 440–447. doi:10.1109/ICICDS70526.2026.11604878
8. N. Bontridder and Y. Poullet, "The role of artificial intelligence in disinformation," *Data Policy*, vol. 3, p. e32, 2021.
9. Performance Analysis of Wireless Sensor Networks for Smart Monitoring Applications. (2016). *International Journal of Humanities and Information Technology, 1*(04), 10–29. doi:10.21590/ijhit.01.04.04
10. A. Vaswani et al., "Attention is all you need," in *Proceedings of Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.
11. Anjani Haritha Sannidhanam. (2023). Zero-Downtime AI Model Updates in Real-Time Inference Systems. *International Journal of Engineering Science & Humanities, 13*(2), 70–78.
12. 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 and Knowledge Discovery*, vol. 10, no. 6, p. e1385, 2020.
13. Vishwanath Nikhil, Dr. Mohammed Abdul Bari, "*Real Time Powerlifting Form Assessment using Yolov5 and Mediapipe*," *International Journal of Information Technology and Computer Engineering*, ISSN 2347–3657, Volume 13, Special Issue 2s, 2025, Pages 574–584.
14. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.
15. Anjani Haritha Sannidhanam. (2024). Prompt Engineering Patterns for Reliable LLM Outputs in Production Environments. *Journal of Multidisciplinary Knowledge, 4*(1), 29–39.
16. 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," *Journal of King Saud University – Computer and Information Sciences*, vol. 35, no. 6, Art. no. 101571, 2023.
17. Sannidhanam, A. H., & Alladi, S. (2017). Cloud-Based Healthcare CRM Systems for Improved Patient Engagement. *International Journal of Technology, Management and Humanities, 3*(01), 32–44. doi:10.21590/ijtmh.3.03.4
18. 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, Art. no. e35865, 2024.
19. L.K. Suresh Kumar, Mohammed Abdul Bari, "*Zero-Day Attack Detection In Multi-Tenant Cloud Environments Using Variational Autoencoders*," *International Journal of Applied Mathematics*, ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version), Volume 38 No. 3s, 2025.
20. A. Orhan, "Fake news detection on social media: The predictive role of university students' critical thinking dispositions and new media literacy," *Smart Learning Environments*, vol. 10, no. 1, p. 29, 2023.
21. R. Zhao and T. Shi, "The impact of large language models on public security

intelligence work and countermeasures research," *Journal of Big Data Computing*, vol. 2, no. 1, pp. 91–103, Mar. 2024.
22. Anjani Haritha Sannidhanam. (2026). LLM-Driven Workflow Optimization: Intelligent Routing in Asynchronous Distributed Systems. *International Journal of AI and Machine Learning, 1*(2), 23–33.
23. 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," *Journal of Experimental Political Science*, vol. 9, no. 1, pp. 104–117, 2022.
24. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali, "*Smartphone Security and Protection Practices*," *International Journal of Engineering and Applied Computer Science (IJEACS)*; ISBN: 9798799755577, Volume: 03, Issue: 01, December 2021 (International Journal, U.K.), Pages 1–6.
25. H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2008.
26. E. Shushkevich, M. Alexandrov, and J. Cardiff, "Improving multiclass classification of fake news using BERT-based models and ChatGPT-augmented data," *Inventions*, vol. 8, no. 5, p. 112, 2023.

About Author:

Anam shanzay is currently pursuing a Master of Technology in Computer Science and Engineering at ISL Engineering College, affiliated to Osmania University, Hyderabad, India. She completed her Bachelor of Engineering in Information Technology from ISL Engineering College, affiliated to Osmania University, from 2019 to 2023.