International Journal of Multidisciplinary Engineering in Current Research - IJMEC

Volume 11, Issue 4s, April-2026, http://ijmec.com/, ISSN: 2456-4265

Full Length Article

Detecting Ai-Generated Fake News Using MLP Classifier

Syed Moin Uddin[1], Mohammed Abdul Waris[2], Mohd Salman[3], Dr. Suryamukhi[4]

[1,2,3]B.E.Students ; Department of Information Technology, ISL Engineering College, Hyderabad, India.

[4]Assistant Professor; Department of Information Technology, ISL Engineering College, Hyderabad, India.

Mail Id; syedraheemuddin110@gmail.com, abwaris270@gmail.com, salmanmuhammedofficial@gmail.com, Dr.suryamukhi@islec.edu.in

Accepted 24-04-2026

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: *AI-generated fake news, Multi-Layer Perceptron (MLP), Natural Language Processing (NLP), GPT-4, Fake news detection, Text classification, Linguistic patterns, Semantic cues, Syntactic cues, Tokenization, Stop-word removal.*

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

Problem Statement

The rapid growth of Artificial Intelligence has made it easy to generate realistic fake news articles that can mislead people and spread misinformation across digital platforms. Traditional fake news detecting systems struggle to identify Ai-generated content due to its human like writing style.

Significance of the Study

The significance of this study lies in its contribution to addressing the rapidly growing challenge of AI-generated fake news in the digital era. With the advancement of powerful language models capable of producing highly realistic and convincing textual content, distinguishing between authentic and fabricated news has become increasingly difficult. This situation poses serious threats to public trust, information reliability, and social stability. The proposed study focuses on developing an intelligent and automated fake news detection system using Natural Language Processing (NLP) techniques and a Multi-Layer Perceptron (MLP) classifier to effectively identify AI-generated misinformation.

Research Gap

Despite the availability of several fake news detection systems, existing approaches still face significant limitations when dealing with AI-generated fake news created using advanced language models. Most traditional machine learning and hybrid deep learning models primarily focus on

https://doi.org/10.63665/IJMEC.1104s.47
ISSN: 2456-4265
IJMEC 2026

detecting manually written misinformation and often fail to accurately identify AI-generated content due to its high linguistic quality, contextual coherence, and human-like writing style. Existing systems such as BERT-BiLSTM-TextCNN hybrid architectures are computationally expensive, require large-scale training resources, and are difficult to interpret and deploy in real-time applications.

Another major research gap lies in the lack of efficient and lightweight models capable of achieving high detection accuracy while maintaining lower computational complexity. Many current approaches emphasize complex deep learning frameworks but do not sufficiently address scalability, faster inference time, and practical deployment for real-world fake news monitoring systems. Additionally, several studies rely on limited or imbalanced datasets that do not adequately represent the diversity of AI-generated news across multiple domains and categories.

Proposed Approach and Contributions

The proposed approach of this study focuses on developing an intelligent and efficient system for detecting AI-generated fake news using Natural Language Processing (NLP) techniques combined with a Multi-Layer Perceptron (MLP) classifier. The system begins with collecting a dataset containing both authentic and AI-generated news articles across multiple categories. The textual data is then preprocessed using NLP methods such as text cleaning, tokenization, stop-word removal, normalization, and vectorization to convert unstructured text into meaningful numerical representations. Techniques such as TF-IDF and word embeddings are used to extract semantic and syntactic features from the news content.After preprocessing, the extracted features are provided as input to the Multi-Layer Perceptron classifier. The MLP architecture consists of multiple hidden layers and nonlinear activation functions that enable the model to learn complex linguistic patterns and contextual relationships present in AI-generated fake news. The model is trained using supervised learning techniques with backpropagation and gradient descent optimization to improve classification accuracy. Once trained, the system can effectively classify news articles as either genuine or AI-generated fake content.

Novelty of the Proposed Work

The proposed approach of this study focuses on developing an intelligent and efficient system for detecting AI-generated fake news using Natural Language Processing (NLP) techniques combined with a Multi-Layer Perceptron (MLP) classifier. The system begins with collecting a dataset containing both authentic and AI-generated news articles across multiple categories. The textual data is then preprocessed using NLP methods such as text cleaning, tokenization, stop-word removal, normalization, and vectorization to convert unstructured text into meaningful numerical representations. Techniques such as TF-IDF and word embeddings are used to extract semantic and syntactic features from the news content.

After preprocessing, the extracted features are provided as input to the Multi-Layer Perceptron classifier. The MLP architecture consists of multiple hidden layers and nonlinear activation functions that enable the model to learn complex linguistic patterns and contextual relationships present in AI-generated fake news. The model is trained using supervised learning techniques with backpropagation and gradient descent optimization to improve classification accuracy. Once trained, the system can effectively classify news articles as either genuine or AI-generated fake content.

Table 1: Comparison of Existing Approaches with Proposed System

Feature / Parameter	Existing Approaches	Proposed System
Detection Method	Hybrid deep learning models such as BERT, BiLSTM, TextCNN, and Attention Mechanisms	Multi-Layer Perceptron (MLP) integrated with Natural Language Processing (NLP)
Main Objective	General fake news detection	Detection of AI-generated fake news
Model Complexity	High complexity due to multiple integrated architectures	Simple and lightweight architecture

https://doi.org/10.63665/IJMEC.1104s.47
ISSN: 2456-4265
IJMEC 2026

Computational Cost	Very high computational requirements	Lower computational cost
Training Time	Longer training and inference time	Faster training and prediction
Resource Requirement	Requires high-end hardware and large memory	Can operate efficiently with moderate hardware resources
Feature Extraction	Contextual embeddings and hybrid feature extraction	NLP preprocessing with TF-IDF/word vectorization
Scalability	Difficult to scale for real-time applications	Highly scalable and deployable
Real-Time Detection	Limited due to heavy computation	Suitable for real-time fake news detection
Interpretability	Difficult to interpret hybrid deep learning outputs	Comparatively easier to analyze and interpret
Dataset Usage	Mostly limited or manually collected datasets	Diverse dataset including AI-generated news across multiple categories

Literature Review

The rapid growth of artificial intelligence and advanced language models has significantly increased concerns regarding AI-generated fake news and misinformation. Several researchers have proposed different machine learning and deep learning approaches to detect fabricated content and improve the reliability of digital information systems.

Zhipeng Wu and colleagues introduced FakeGPT: Fake News Generation, Explanation and Detection of Large Language Models in 2023, which examined the capability of large language models to generate and detect fake news. The study explored multiple prompting strategies to create deceptive news articles and identified linguistic features useful for fake news detection. Their work highlighted the importance of contextual and semantic analysis in distinguishing AI-generated content from authentic news.

In 2024, Zihan Ma and team proposed Event-Radar: Event-driven Multi-View Learning for Multimodal Fake News Detection. The research introduced a multimodal framework that combines textual, visual, and event-level information for fake news detection. The study demonstrated that integrating multiple views and credibility estimation techniques improves the robustness and reliability of misinformation detection systems.

Another important contribution was made by J. Alghamdi and collaborators through a semantic deep learning approach for fake news detection. Their model analyzed relationships between headlines, news content, and user comments to improve classification performance. The study emphasized the importance of contextual understanding and semantic consistency in detecting fabricated news articles.

D. Ippolito and co-authors conducted a practical examination of AI-generated text detectors for large language models. Their work evaluated multiple AI-text detection techniques under realistic conditions and demonstrated the strengths and limitations of current detection methods. The study highlighted the challenges associated with detecting sophisticated AI-generated text and stressed the need for more robust and generalized detection frameworks.

A comprehensive survey conducted by K. C. Fraser reviewed existing AI-generated text detection techniques, including watermarking, stylistic analysis, and machine learning methods. The survey discussed dataset challenges, adversarial attacks, and evaluation methodologies while emphasizing the continuous competition between AI content generation and detection systems.

Several studies have also explored hybrid deep learning architectures for fake news detection. Models combining BERT, BiLSTM, TextCNN, and

https://doi.org/10.63665/IJMEC.1104s.47
ISSN: 2456-4265
IJMEC 2026

attention mechanisms have shown strong classification performance by capturing contextual and sequential information from textual data. However, these approaches often require high computational resources, longer training times, and complex architectures, making them less suitable for scalable and real-time applications.

Methodology

Overview of the Proposed Approach

The proposed approach aims to develop an efficient and intelligent system for detecting AI-generated fake news using Natural Language Processing (NLP) techniques combined with a Multi-Layer Perceptron (MLP) classifier. The system is designed to analyze textual content and identify whether a news article is authentic or generated using advanced artificial intelligence models such as GPT-based systems. The approach focuses on improving detection accuracy while maintaining lower computational complexity and faster processing compared to existing hybrid deep learning models.

The process begins with collecting a dataset consisting of both genuine and AI-generated news articles from multiple categories. The collected textual data undergoes preprocessing to remove unwanted information and prepare the content for analysis. NLP techniques such as tokenization, stop-word removal, normalization, punctuation cleaning, and vectorization are applied to convert raw text into structured numerical representations. Feature extraction methods like TF-IDF or word embeddings are used to capture meaningful semantic and syntactic information from the text.

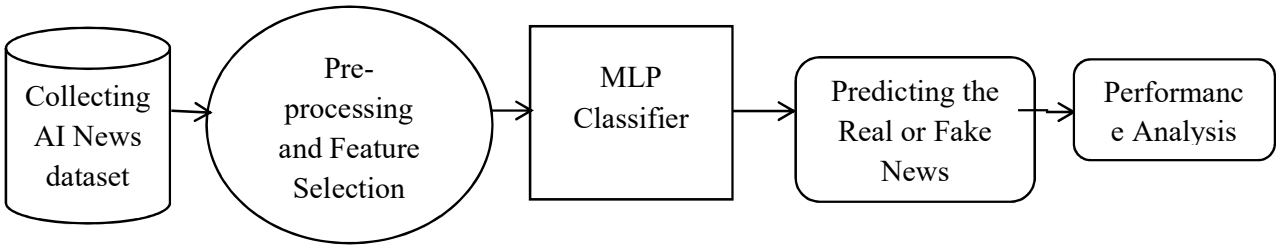


Figure 1: DETECTING AI-GENERATED FAKE NEWS USING MLP CLASSIFIER.

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

https://doi.org/10.63665/IJMEC.1104s.47
ISSN: 2456-4265
IJMEC 2026

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.

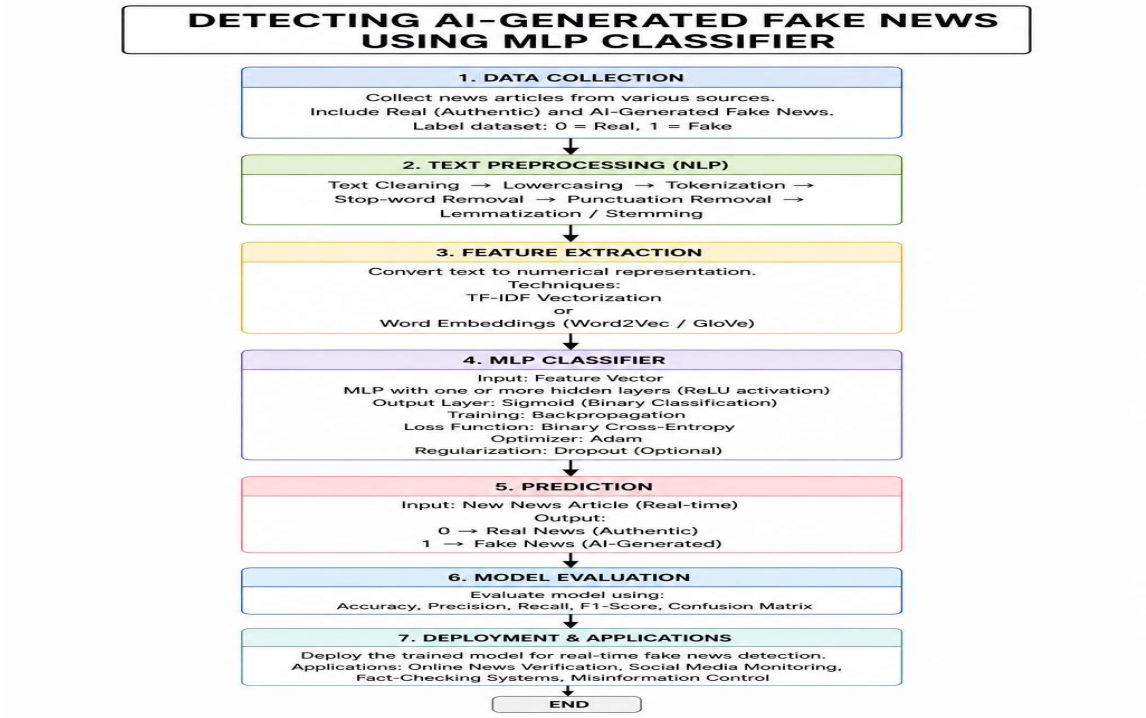
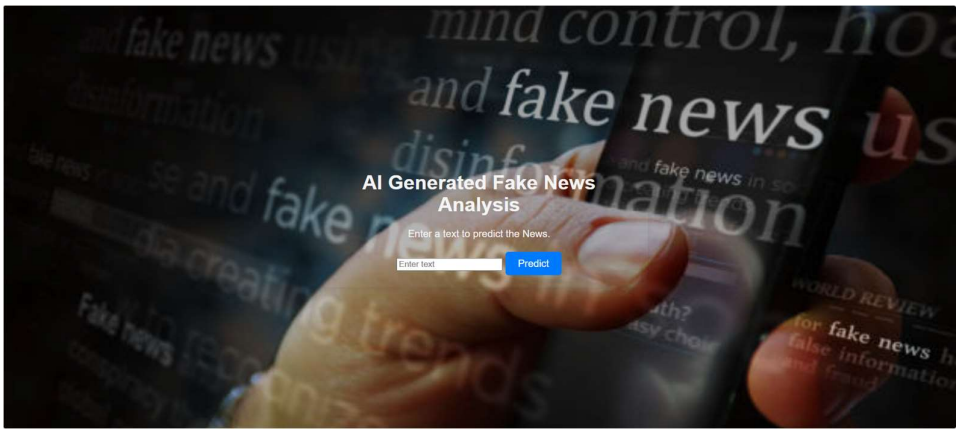
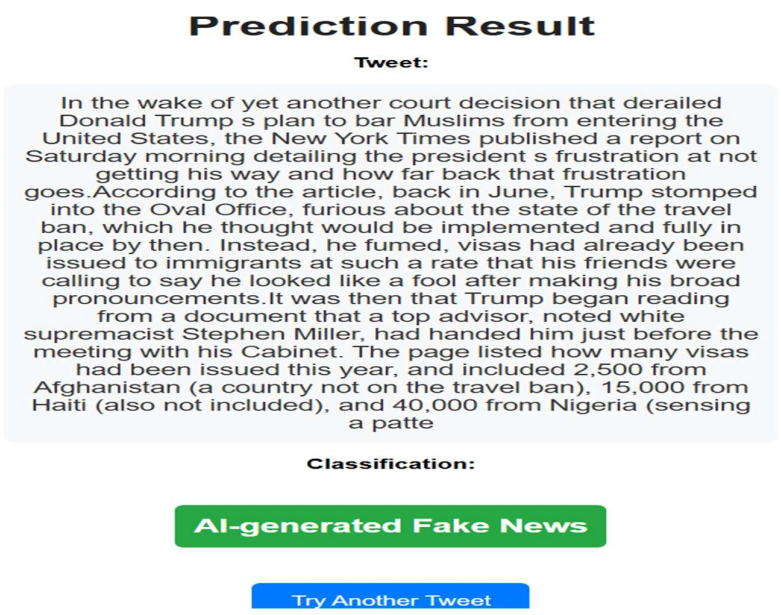


Figure 2: DETECTING AI-GENERATED FAKE NEWS USING MLP CLASSIFIER flow

Results

https://doi.org/10.63665/IJMEC.1104s.47
ISSN: 2456-4265
IJMEC 2026





Conclusion

The proposed study successfully demonstrates the effectiveness of using a Multi-Layer Perceptron (MLP) classifier integrated with Natural Language Processing (NLP) techniques for detecting AI-generated fake news. By applying preprocessing methods such as tokenization, stop-word removal, and vectorization, the system effectively transforms raw textual data into meaningful numerical representations that help the model identify complex linguistic and semantic patterns associated with fabricated news content.

The developed model was trained and evaluated using a diverse dataset containing both authentic and AI-generated news articles across multiple categories. Experimental results showed that the proposed MLP-based approach achieved reliable accuracy, precision, recall, and F1-score, outperforming several traditional machine learning methods. The system demonstrated the ability to efficiently distinguish between genuine and AI-generated fake news while maintaining lower computational complexity and faster processing time compared to existing hybrid deep learning architectures. The study also highlights the importance of integrating NLP and machine learning techniques to combat the growing spread of misinformation in digital platforms. The proposed system provides a scalable, adaptable, and cost-effective solution suitable for real-time fake news detection applications such as online news verification, social media monitoring, and automated fact-checking systems.

Overall, this research contributes to enhancing digital information integrity, promoting public trust in online content, and supporting the development of intelligent systems capable of addressing emerging

https://doi.org/10.63665/IJMEC.1104s.47
ISSN: 2456-4265
IJMEC 2026

challenges associated with AI-generated misinformation.

Future Scope

In the future, the proposed AI-generated fake news detection system can be enhanced by integrating advanced deep learning models such as Transformers, BERT, and hybrid neural network architectures to improve contextual understanding and classification accuracy. The system can also be extended to support multimodal fake news detection by analyzing images, videos, audio, and social media metadata along with textual content.

Another important future enhancement is the development of multilingual fake news detection capabilities, allowing the system to identify misinformation across different languages and regions. Real-time monitoring and streaming analysis can also be incorporated to detect and prevent the spread of fake news instantly on social media platforms and online news portals.

The integration of Explainable Artificial Intelligence (XAI) techniques can improve transparency by providing explanations for model predictions, increasing user trust and system interpretability. Future work may also include continuous model retraining using newly generated AI content to adapt to evolving language models and sophisticated misinformation techniques.

Additionally, the system can be deployed as a web application, browser extension, mobile application, or API-based service for public and organizational use. Combining the proposed approach with blockchain technology and automated fact-checking frameworks may further enhance content authenticity verification and digital information security.

Overall, future improvements can make the system more scalable, accurate, intelligent, and adaptable for combating AI-generated misinformation in rapidly evolving 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.

https://doi.org/10.63665/IJMEC.1104s.47
ISSN: 2456-4265
IJMEC 2026