TruthGuard AI: A Hybrid Fake News Detection Platform Using Rule-Based Analysis, Machine Learning, and LLM Verification

PICHIKALA ASHA JYOTHI

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

K. Venkatesh

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The rapid proliferation of digital media platforms has significantly increased the spread of misinformation and fake news, posing serious threats to public trust, democratic processes, and societal stability. Traditional methods of detecting fake news often rely on either manual verification or single-model automated approaches, which are limited in scalability, accuracy, and contextual understanding. To address these challenges, this project introduces **TruthGuard AI**, a hybrid fake news detection platform that integrates rule-based linguistic analysis, machine learning inference using TF-IDF, and a Large Language Model (LLM) verification layer.The proposed system operates in a multi-stage pipeline designed to enhance both efficiency and reliability. In the first stage, rule-based analysis evaluates textual content for clickbait patterns, sensational language, and known fake-news indicators such as exaggerated claims or conspiracy-related keywords. This stage provides a quick heuristic filter that identifies suspicious linguistic features. In the second stage, the system applies a TF-IDF-based machine learning model to analyze term importance and contextual relevance, enabling statistical classification of news as credible or fake. This approach improves detection accuracy by leveraging historical patterns in labeled datasets.The third stage introduces an advanced verification layer powered by a local LLM (via Ollama), which performs deep contextual reasoning. The LLM analyzes the content for bias, missing sources, logical inconsistencies, and credibility signals, offering an explainable and human-like interpretation of the news. This hybrid architecture ensures that the limitations of one method are compensated by the strengths of others, resulting in a more robust detection system.The platform is implemented as a user-friendly desktop application using Python's Tkinter and ttkbootstrap libraries, providing an interactive interface for users to input news articles or headlines. The system outputs a reliability score, a verdict (credible, suspicious, or fake), and a detailed reasoning log that enhances transparency and user trust.Experimental results demonstrate that combining heuristic, statistical, and deep learning-based approaches significantly improves detection performance compared to standalone methods. The system also supports scalability and extensibility, allowing integration with real-time news feeds and cloud-based LLM services in future enhancements.In conclusion, TruthGuard AI represents a comprehensive and practical solution for combating misinformation. By combining traditional and modern AI techniques, it not only detects fake news effectively but also provides explainable insights, making it suitable for real-world deployment in journalism, education, and social media monitoring.

Keywords: Fake News Detection, Hybrid AI, TF-IDF, Natural Language Processing, Machine Learning, Large Language Models, Rule-Based Systems, Misinformation, Text Classification, Explainable AI

I. INTRODUCTION

In the digital age, the consumption of news has shifted dramatically from traditional media outlets to online platforms such as social media, blogs, and independent news websites. While this transition has democratized information sharing, it has also led to the widespread dissemination of fake news—false or misleading information presented as legitimate news. The impact of fake news is profound, influencing public opinion, inciting social unrest, and even affecting political outcomes.Detecting fake news is a complex challenge due to the nuanced nature of language, the intentional design of misleading content, and the rapid speed at which information spreads online. Early approaches to fake news detection relied heavily on manual fact-checking, which is time-consuming and not scalable. With advancements in artificial intelligence and natural language processing (NLP), automated systems have been developed to classify news articles based on linguistic and statistical features. However, these systems often struggle with contextual understanding and adaptability.This project proposes a hybrid approach to fake news detection, combining multiple methodologies to improve accuracy and reliability. The system integrates three key components: rule-based analysis, machine learning using TF-IDF, and Large Language Model (LLM) verification. Each component addresses specific limitations of the others, creating a balanced and efficient detection mechanism.Rule-based analysis focuses on identifying linguistic patterns commonly associated with fake news, such as sensational phrases, emotional triggers, and conspiracy-related terminology. While this method is fast and interpretable, it lacks the ability to generalize beyond predefined rules. To overcome this, the system incorporates a TF-IDF-based machine learning model, which statistically evaluates the importance of words in a document and uses this information for classification. This approach enhances the system's ability to learn from data and adapt to new patterns.The final layer involves the use of a Large Language Model, which performs deep semantic analysis of the content. Unlike traditional models, LLMs can understand context, detect subtle biases, and provide detailed explanations for their conclusions. By integrating this layer, the system achieves a higher level of interpretability and accuracy.The implementation of this system as a desktop application ensures accessibility and ease of use. Users can input news content and receive a reliability score along with a detailed explanation of the analysis process. This transparency is crucial for building trust in automated systems.Overall, this project aims to provide a comprehensive solution to the fake news problem by leveraging the strengths of different AI techniques. It highlights the importance of hybrid models in addressing complex real-world challenges and sets the foundation for future research in misinformation detection.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

The detection of fake news has been an active area of research, particularly with the rise of social media platforms. Various approaches have been proposed, ranging from traditional machine learning techniques to advanced deep learning models.Early studies focused on **rule-based systems**, where predefined linguistic patterns and keywords were used to identify fake news. These systems relied on detecting sensational phrases, excessive punctuation, and emotionally charged language. While effective for simple cases, rule-based approaches lack scalability and fail to capture complex contextual relationships.Subsequently, researchers explored **machine learning techniques**, including Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression. These models often utilized features such as word frequency, n-grams, and TF-IDF vectors to classify news articles. For instance, TF-IDF became a widely used method due to its ability to quantify the importance of words in a document relative to a corpus. Although these methods improved accuracy, they required extensive feature engineering and struggled with semantic understanding.With advancements in deep learning, **neural network-based models** such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) were introduced. These models automatically extract features from text and can capture sequential dependencies. However, they require large datasets and significant computational resources, making them less practical for real-time applications.More recently, **transformer-based models** and Large Language Models (LLMs) have revolutionized NLP tasks. Models like BERT and GPT can understand context, detect subtle nuances, and generate human-like explanations. Researchers have used these models for fake news detection by fine-tuning them on labeled datasets. While highly accurate, these models can be resource-intensive and may lack transparency in decision-making.Hybrid approaches have also been proposed, combining multiple techniques to leverage their strengths. For example, some systems integrate rule-based filtering with machine learning classifiers, while others incorporate knowledge graphs or external fact-checking APIs. These approaches have shown improved performance but often lack a unified framework that balances speed, accuracy, and interpretability.The proposed system builds upon these advancements by integrating rule-based analysis, TF-IDF-based machine learning, and LLM verification into a single pipeline. This hybrid architecture addresses the limitations of individual methods and provides a more robust solution for fake news detection.

III. EXISTING SYSTEM

Existing fake news detection systems primarily rely on either manual verification or single-method automated approaches. Manual fact-checking, conducted by organizations and journalists, ensures high accuracy but is time-consuming and not scalable for the vast amount of online content generated.Automated systems often use **machine learning models** trained on labeled datasets. These systems analyze textual features such as word frequency, sentence structure, and sentiment to classify news articles. While effective to some extent, they are limited by their dependence on training data and inability to adapt to new forms of misinformation.

Another category includes **rule-based systems**, which detect fake news using predefined keywords and patterns. These systems are fast and interpretable but lack flexibility and fail to handle complex linguistic variations. They are also prone to high false positives when legitimate content contains similar patterns.More advanced systems utilize **deep learning and transformer models**, which offer improved accuracy and contextual understanding. However, these models require significant computational resources and are often deployed in cloud environments, making them less accessible for standalone applications. Additionally, they may lack transparency, making it difficult for users to understand how decisions are made.Overall, existing systems face challenges in balancing accuracy, scalability, interpretability, and computational efficiency. This highlights the need for a hybrid approach that combines multiple techniques to overcome these limitations, which is the primary motivation behind the proposed TruthGuard AI system.

IV. PROPOSED METHOD

The proposed system, **TruthGuard AI**, is a hybrid fake news detection platform designed to overcome the limitations of traditional single-method approaches. It integrates three core components—rule-based linguistic analysis, TF-IDF-based machine learning inference, and a Large Language Model (LLM) verification layer—into a unified multi-stage pipeline.The first stage performs **rule-based analysis**, where the input news content is evaluated for predefined linguistic patterns commonly associated with fake news. These include sensational phrases, exaggerated claims, and conspiracy-related keywords. This stage provides a rapid preliminary assessment and filters out highly suspicious content with minimal computational overhead.In the second stage, the system applies **TF-IDF (Term Frequency–Inverse Document Frequency)** to convert textual data into numerical feature vectors. These vectors are then processed using machine learning classifiers such as Logistic Regression or Support Vector Machines to determine the statistical likelihood of the content being fake or credible. Studies show that TF-IDF-based models remain effective for text classification due to their ability to capture term importance and contextual relevance .The third stage introduces an **LLM-based verification layer** using a local model (via Ollama). This layer performs deep semantic analysis, identifying bias, missing sources, logical inconsistencies, and factual gaps. Unlike traditional models, LLMs provide explainable reasoning, enhancing transparency and trust in the system. Recent research highlights the importance of LLMs in generating interpretable justifications for fake news detection .The final output combines results from all three stages to generate a **reliability score**, a **verdict (credible, suspicious, or fake)**, and a **detailed reasoning log**. This hybrid architecture ensures improved accuracy, robustness, and explainability, making the system suitable for real-world deployment.

V. IMPLEMENTATION

The implementation of TruthGuard AI is carried out using Python, integrating multiple libraries for GUI development, text processing, and API communication. The system is designed as a desktop application to ensure accessibility and ease of use.The **front-end interface** is developed using Tkinter and ttkbootstrap, which provides a modern and responsive user interface. The GUI includes input fields for entering news content, control buttons for initiating analysis, and output sections displaying reliability scores, verdicts, and reasoning logs. The use of a progress bar and threading ensures that the interface remains responsive during computation.The **backend processing pipeline** is divided into three major stages. In the first stage, rule-based analysis is implemented using Python's string processing and regular expressions. A predefined list of fake and real keywords is used to compute an initial heuristic score. This stage is lightweight and provides instant feedback.In the second stage, the system incorporates **TF-IDF vectorization**, typically implemented using libraries such as Scikit-learn. The input text is preprocessed through tokenization, stop-word removal, and normalization. The TF-IDF vectorizer transforms the text into numerical features, which are then fed into a trained machine learning model. Research indicates that TF-IDF combined with classifiers like Logistic Regression and SVM achieves high accuracy in fake news detection tasks .The third stage involves integration with a **local LLM via Ollama API**. The system sends a prompt containing the news content to the LLM, requesting an analysis of bias, credibility, and missing evidence. The response is processed and displayed as part of the reasoning log. This stage enhances the system's ability to perform contextual and semantic evaluation.To ensure smooth execution, **multithreading** is used to handle long-running tasks such as API calls, preventing the GUI from freezing. Error handling mechanisms are also implemented to manage cases where the LLM is unavailable, allowing the system to fall back to heuristic analysis.The final results are aggregated and presented to the user in a clear and interpretable format. The modular design of the system allows for future enhancements, such as integrating real-time news APIs, cloud-based LLMs, or advanced deep learning models.

VI. ALGORITHMS

The proposed system utilizes a combination of rule-based, statistical, and AI-driven algorithms to detect fake news effectively.

1. Rule-Based Scoring Algorithm

This algorithm scans the input text for predefined keywords associated with fake or credible news. Each occurrence contributes to a weighted score:

- Fake keywords reduce the score
- Real keywords increase the score

Formula:
Score = 50 + (Real_hits × 10) − (Fake_hits × 15)

The score is normalized between 0 and 100. This method provides a fast and interpretable baseline.

2. TF-IDF Vectorization Algorithm

TF-IDF converts textual data into numerical form by evaluating:

- **Term Frequency (TF):** Frequency of a word in a document
- **Inverse Document Frequency (IDF):** Importance of a word across documents

Formula:
TF-IDF = TF × log(N / DF)

Where N is total documents and DF is document frequency. This approach helps highlight significant words while reducing noise from common terms. TF-IDF is widely used in fake news detection for feature extraction .

3. Machine Learning Classification

The TF-IDF vectors are fed into classifiers such as:

- Logistic Regression
- Support Vector Machine (SVM)

These models learn patterns from labeled datasets and classify new inputs accordingly.

4. LLM-Based Verification Algorithm

The LLM processes the input text using prompt-based inference. It evaluates:

- Source credibility
- Logical consistency
- Bias detection
- Missing evidence

It then generates a natural language explanation, improving interpretability. Recent studies show that LLM-based approaches enhance robustness and reasoning capabilities in fake news detection .

VII. SYSTEM DESIGN

The system design of TruthGuard AI follows a modular and layered architecture to ensure scalability, maintainability, and efficiency. The architecture is divided into four primary components: User Interface Layer, Processing Layer, Intelligence Layer, and Output Layer.

1. User Interface Layer

This layer is responsible for user interaction. It is implemented using Tkinter and ttkbootstrap, providing a visually appealing and responsive interface. Users can input news content, initiate analysis, and view results. The interface includes text areas, buttons, progress indicators, and result panels.

2. Processing Layer

The processing layer handles input preprocessing and workflow management. It performs tasks such as:

- Text cleaning (removing special characters, stop words)
- Tokenization and normalization
- Managing asynchronous execution using threading

This layer ensures that the system processes data efficiently without affecting the user experience.

3. Intelligence Layer

This is the core component of the system, consisting of three submodules:

a) Rule-Based Engine

This module applies predefined linguistic rules to detect suspicious patterns. It acts as a fast filtering mechanism.

b) Machine Learning Engine

This module uses TF-IDF vectorization and classification algorithms to analyze text statistically. It leverages trained models to predict the authenticity of news content. Research shows that hybrid NLP and ML frameworks significantly improve detection accuracy .

c) LLM Verification Engine

This module integrates a local LLM via API. It performs deep contextual analysis and generates explanations. It enhances interpretability and reduces false positives.

4. Output Layer

The output layer aggregates results from all modules and presents them to the user. It includes:

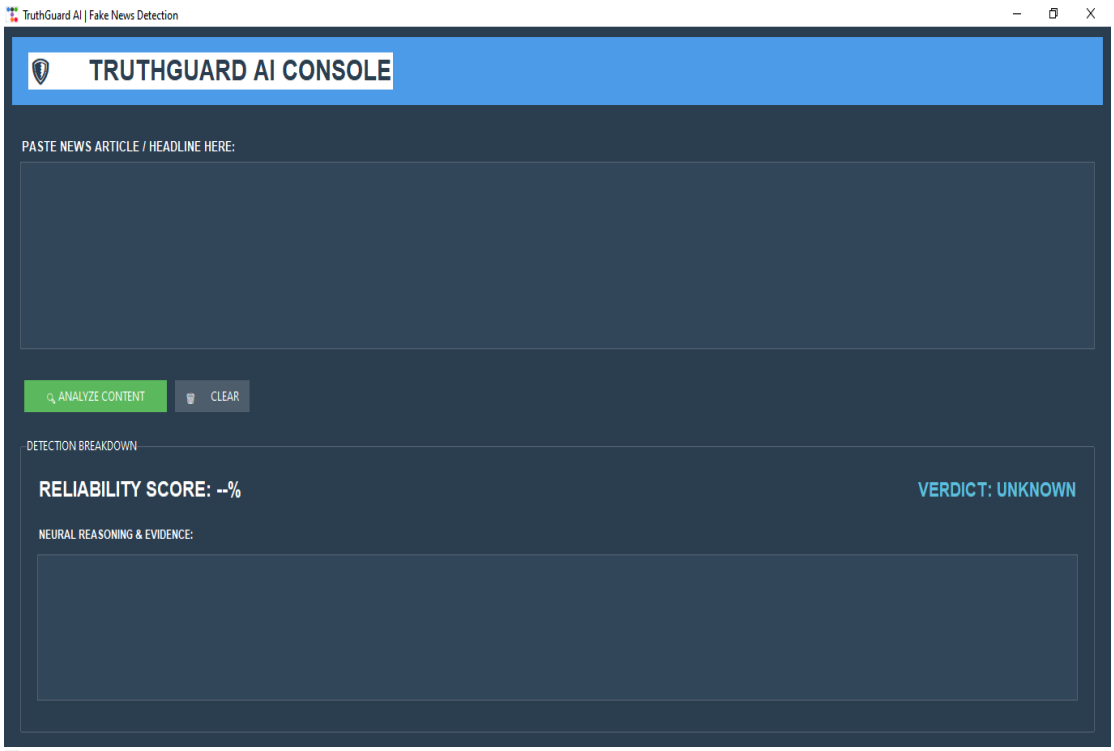
- Reliability score (0–100%)
- Final verdict (credible, suspicious, fake)
- Detailed reasoning log

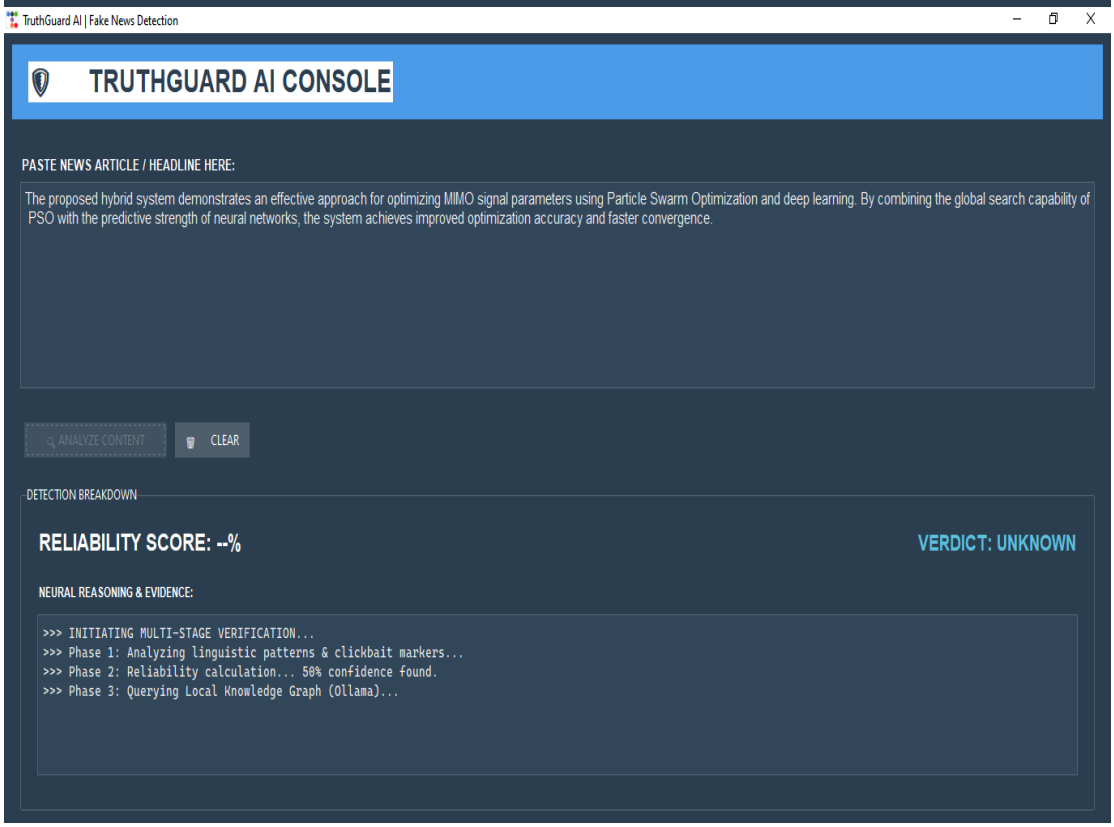
Data Flow

1. User inputs news text
2. Text is preprocessed
3. Rule-based analysis computes initial score
4. TF-IDF model performs classification
5. LLM generates contextual insights
6. Results are combined and displayed

This modular design ensures flexibility, allowing easy integration of additional components such as real-time fact-checking APIs or multi-modal analysis systems. Modern architectures increasingly adopt hybrid frameworks to balance speed, accuracy, and interpretability in fake news detection systems .

SYSTEM DESIGN IMAGES





VIII. CONCLUSION

The increasing spread of fake news in the digital era necessitates robust and scalable detection systems. This project presents **TruthGuard AI**, a hybrid fake news detection platform that combines rule-based analysis, machine learning, and LLM-based verification to address the limitations of existing approaches.The system demonstrates that integrating multiple techniques significantly enhances detection accuracy and reliability. The rule-based component provides fast initial screening, while the TF-IDF-based machine learning model offers statistical validation. The inclusion of an LLM layer enables deep contextual understanding and explainable reasoning, which are critical for building user trust.One of the key strengths of the proposed system is its **hybrid architecture**, which balances performance, interpretability, and computational efficiency. Unlike traditional systems that rely on a single method, this approach leverages the strengths of each technique to produce more accurate and meaningful results. Additionally, the user-friendly interface ensures accessibility for non-technical users.The system is designed with scalability in mind, allowing future enhancements such as integration with real-time news APIs, cloud-based LLM services, and multi-modal analysis. Furthermore, it can be extended to support multiple languages and domains, making it applicable in diverse real-world scenarios.In conclusion, TruthGuard AI provides an effective and practical solution for fake news detection. By combining traditional and modern AI techniques, it not only identifies misinformation but also explains the reasoning behind its decisions. This makes it a valuable tool for combating misinformation and promoting information integrity in today's digital landscape.

1. **REFERENCES**Roumeliotis, K. et al. (2025). *Fake News Detection using CNNs and LLMs*. Future Internet.
2. Al-Tarawneh, M. et al. (2024). *Word Embedding Techniques for Fake News Detection*. Computers Journal.
3. Ayyasamy, R. et al. (2025). *Hybrid Deep Learning Framework for Fake News Detection*. Scientific Reports.
4. Khan, Z. (2023). *TF-IDF Weighted Fake News Detection*. IJISAE.
5. Akre, H. (2025). *Fake News Detection using NLP and ML*. ET Journal.
6. Puri, R. (2024). *Comprehensive Study on Fake News Detection*. ResearchGate.
7. Nadeem, M. et al. (2024). *Hybrid NLP-ML Framework*. ICCK Journal.
8. ITDW Journal (2024). *Text Representation Techniques in Fake News Detection*.
9. ScienceDirect (2022). *Linguistic Feature-Based Detection*.
10. Singhal, D. (2024). *TF-IDF and Count Vectorizer Models*. IJESDF.
11. Raza, S. et al. (2024). *BERT vs LLM Fake News Detection*. arXiv.
12. Su, J. et al. (2023). *Bias in Fake News Detection Models*. arXiv.
13. Zhou, Y. et al. (2023). *Multi-modal Fake News Detection*. arXiv.
14. Wang, B. et al. (2024). *Explainable Fake News Detection using LLMs*. arXiv.
15. NAACL (2024). *News Source Reliability Estimation*.