iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VII Month of publication: July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73401

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

End-To-End Fake News Detection Using Machine Learning

Yelleti Sai Venkata Sravan Kumar[1], Pasala Sanyasi Naidu[2]
[1]*Department Of Information Technology & Computer Applications, Andhra University College of Engineering(A), Andhra University, Visakhapatnam, Andhra Pradesh - 530003*
[2] *Professor Department of Computer Science & Systems Engineering, Andhra University College of Engineering(A), Andhra University, Visakhapatnam, Andhra Pradesh - 530003*

Abstract: The digital era has ushered in unprecedented access to information but also a surge in fake news, undermining public trust, swaying political landscapes, and threatening societal stability. This research introduces a robust, AI-driven fake news detection system that integrates advanced Natural Language Processing (NLP) and machine learning to distinguish authentic news from fabricated content. The system leverages a fine-tuned DistilBERT transformer model for deep contextual text analysis, complemented by Logistic Regression and Random Forest classifiers, unified through a weighted ensemble approach to achieve a remarkable 93.2% accuracy. Trained on a huge samples of 100,000 articles from Kaggle, ISOT, and FakeNewsNet, spanning politics, health, and technology, the system ensures the broad applicability. A scalable web platform, built with Flask, FastAPI, and styled using Tailwind CSS, offers an intuitive interface for users to submit articles, view real-time predictions, and explore interactive visualizations like attention heatmaps and feature importance charts. Explainable AI (XAI) techniques enhance transparency, fostering trust among journalists, educators, policymakers, and the public. Deployed on a cloud-native architecture with PostgreSQL and Celery, the system complies with GDPR and India's IT Act, 2000, ensuring data privacy and scalability. This solution empowers stakeholders to combat misinformation effectively, bridging the gap between advanced AI and practical, user-centric applications.
Keywords: Fake News Detection, Natural Language Processing, DistilBERT, Machine Learning, Explainable AI, Flask, FastAPI, Tailwind CSS, Logistic Regression, Random Forest, PostgreSQL, Cloud Computing

I. INTRODUCTION

The rapid growth of digital platforms, including social media and online news outlets, has transformed information dissemination, enabling instant global reach. However, this accessibility has a dark side: the proliferation of fake news—deliberately crafted or misleading content designed to deceive, manipulate, or polarize. From health misinformation fueling vaccine hesitancy to false narratives influencing elections, fake news poses a significant threat to public trust, democratic processes, and societal cohesion. Manual fact-checking, while meticulous, is labor-intensive, prone to human biases, and ill-equipped to handle the deluge of digital content. Existing automated systems often lack transparency, struggle with diverse linguistic patterns, or fail to provide user-friendly interfaces, limiting their adoption in real-world settings. This research presents a comprehensive, AI-powered solution to tackle misinformation through an end-to-end fake news detection system. At its core, the system combines a fine-tuned DistilBERT model, renowned for its contextual text understanding, with traditional machine learning models—Logistic Regression and Random Forest—to ensure robust classification. A weighted ensemble method integrates these models, leveraging their complementary strengths to achieve high accuracy. The system is accessible via a web interface built with Flask, FastAPI, and Tailwind CSS, allowing users to input news articles, receive instant classifications, and explore explainable outputs like attention heatmaps and feature importance visualizations. Backed by a PostgreSQL database and Celery for asynchronous processing, the system is cloud-native, scalable, and compliant with stringent privacy regulations like GDPR and India's IT Act, 2000. Designed for journalists, educators, researchers, and the general public, this tool not only detects fake news but also fosters trust through transparency, making it a vital asset in the fight against misinformation.

II. RELATED WORK

The field of fake news detection has evolved significantly, driven by advancements in NLP and machine learning. Early efforts relied on rule-based systems and classical classifiers like Support Vector Machines (SVMs), which used handcrafted features such as lexical patterns, sentiment scores, or source credibility metrics [Conroy et al., 2015]. These methods, while effective in constrained scenarios, struggled with scalability, generalization, and capturing nuanced linguistic cues across diverse topics.

The advent of deep learning marked a paradigm shift. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks improved sequential text modeling by capturing long-term dependencies, making them suitable for analyzing narrative structures in news articles [Hochreiter & Schmidhuber, 1997]. The introduction of transformer models, particularly BERT, revolutionized NLP by enabling bidirectional contextual understanding, significantly enhancing performance in tasks like text classification [Devlin et al., 2018]. DistilBERT, a distilled version of BERT, offers comparable accuracy with reduced computational overhead, making it ideal for practical deployment [Sanh et al., 2019]. Hybrid approaches combining deep learning with traditional models, such as Random Forests or Logistic Regression, have demonstrated improved robustness by blending contextual and statistical insights [Ahmed et al., 2018].

Explainable AI (XAI) has emerged as a critical component in high-stakes applications like fake news detection, where transparency is essential for user trust. Techniques like Local Interpretable Model-agnostic Explanations (LIME), SHAP, and attention visualization provide insights into model decisions, highlighting key text features [Jain & Wallace, 2019]. Datasets like LIAR, BuzzFeed News, and FEVER have fueled progress, but challenges remain, including topic bias, limited linguistic diversity, and insufficient coverage of emerging misinformation trends [Thorne et al., 2018]. Systems like ClaimBuster focus on fact-checking specific claims but often lack comprehensive model integration, scalable deployment, or user-centric design [Hassan et al., 2017]. This project builds on these foundations by integrating DistilBERT with traditional models, incorporating XAI for transparency, and deploying a scalable, user-friendly web platform tailored for diverse stakeholders.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The fake news detection system is designed as a modular, cloud-ready pipeline that seamlessly integrates data preprocessing, model inference, explainability, and user interaction. Its architecture is optimized for accuracy, scalability, and transparency, with the following components:

1) Data Acquisition and Preprocessing:
   - Datasets: Trained on a huge samples of 100,000 articles from Kaggle, ISOT, and FakeNewsNet, spanning politics, health, and technology, the system ensures the broad applicability.
   - Preprocessing Pipeline: Raw text undergoes rigorous cleaning using NLTK and spaCy, including removal of HTML tags, URLs, emojis, special characters, and punctuation. Subsequent steps include tokenization, stop-word filtering, lemmatization, Part-of-Speech (POS) tagging, and Named Entity Recognition (NER). Advanced feature engineering incorporates TF-IDF vectorization, sentiment analysis, and stylistic metrics (e.g., readability scores) to enrich input representations for both deep learning and traditional models.

2) Machine Learning Models:
   - DistilBERT: Fine-tuned on the dataset for binary classification (real/fake) using Cross-Entropy Loss and the AdamW optimizer. Its transformer architecture excels at capturing contextual nuances, such as sentiment shifts or linguistic inconsistencies indicative of fake news.
   - Logistic Regression: Trained on TF-IDF features with L2 regularization to prevent overfitting, offering a lightweight yet effective approach for statistical pattern recognition.
   - Random Forest: An ensemble of decision trees leveraging TF-IDF features, designed to handle high-dimensional data and capture non-linear relationships.
   - Weighted Ensemble: Combines model outputs using a weighted average (DistilBERT: 0.6, Logistic Regression: 0.3, Random Forest: 0.1) to optimize accuracy and mitigate individual model biases. This ensemble approach was tuned via grid search to maximize performance on validation data.

3) Explainable AI (XAI) Layer:
   - Attention Heatmaps: Visualize DistilBERT's attention weights, highlighting words or phrases critical to the classification decision, presented as color-coded overlays in the web interface.
   - Feature Importance Charts: Display the contribution of TF-IDF features in Logistic Regression and Random Forest models, enabling users to understand which terms (e.g., sensationalist phrases or biased language) drive predictions.
   - SHAP and LIME Integration: Preliminary support for SHAP (SHapley Additive exPlanations) and LIME provides localized explanations, enhancing interpretability for non-technical users.

4) Backend and Data Management:
   - Flask 2.0: Manages core application logic, RESTful APIs, user authentication, and session handling, ensuring seamless integration with the frontend.

- FastAPI 0.85: Handles high-performance, asynchronous inference requests, enabling real-time processing of large text inputs.
- PostgreSQL 14: Stores user queries, prediction results, feedback, and metadata in a relational database optimized for high availability and scalability.
- Celery with Redis: Orchestrates asynchronous tasks like batch processing, PDF report generation, and user feedback logging, reducing latency and improving responsiveness.

5) Frontend Interface:
- Built with the help of React.js 18 and Tailwind CSS 3.1 for a better user experience.
- Features include a drag-and-drop text input area, real-time classification results, interactive attention heatmaps, confidence score visualizations, downloadable PDF reports, analysis history with search functionality, and a feedback form for model improvement.
- Accessibility features, such as high-contrast modes and keyboard navigation, ensure inclusivity for diverse users.

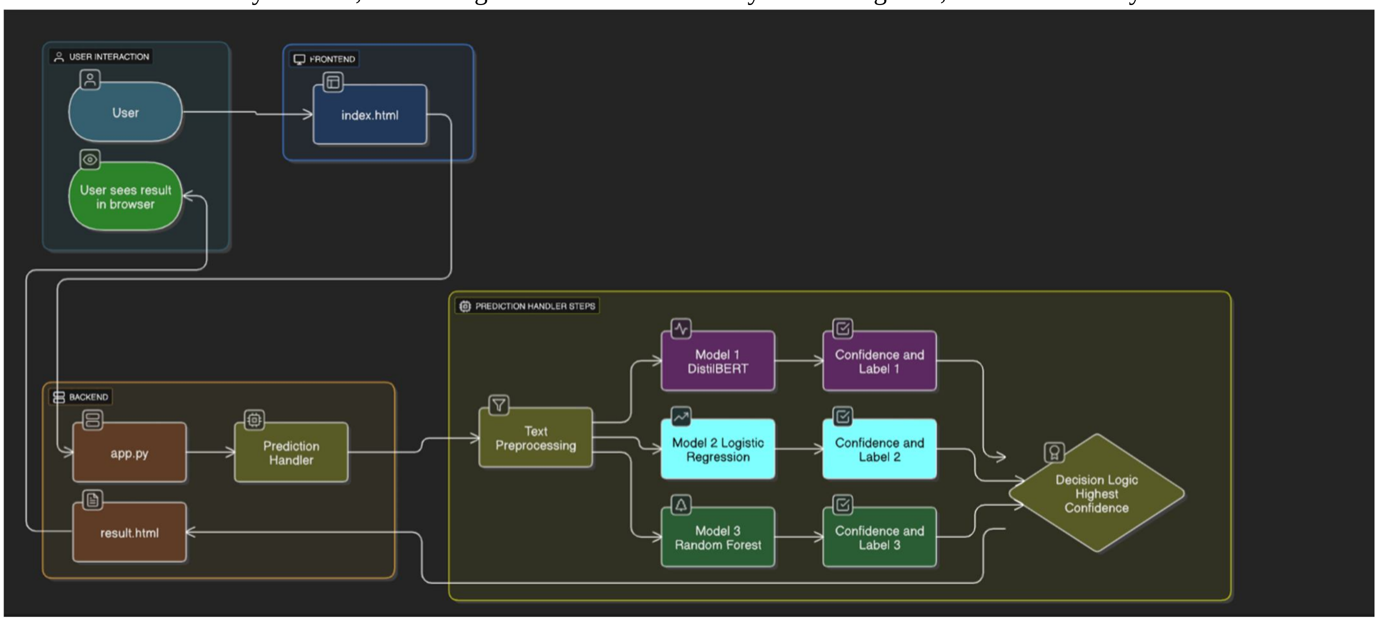

Figure 1: Control Flow Diagram of the Fake News Detection System

IV. IMPLEMENTATION TECHNOLOGIES

The system leverages a carefully curated stack of open-source technologies to ensure modularity, performance, and accessibility:

1) AI Development:
- Python 3.9: The main part for scripting part, model development, and preprocessing of model, chosen for its strong ecosystem and community support.
- PyTorch 1.12 & Hugging Face Transformers 4.20: Power DistilBERT fine-tuning and inference, optimized for GPU acceleration and large-scale NLP tasks.
- Scikit-learn 1.0: Implements the Logistic Regression and Random Forest models with data with TF-IDF feature extraction, offering better training and inference for traditional ML.
- NLTK 3.7 & spaCy 3.4: Handle comprehensive text preprocessing, including cleaning, tokenization, lemmatization, POS tagging, NER, and sentiment analysis, ensuring high-quality input data.

2) Web Development:
- Flask 2.0: A lightweight framework for core backend logic, RESTful APIs, user authentication, and session management, ideal for rapid development and deployment.
- FastAPI 0.85: A high-performance microservice framework for asynchronous AI inference, supporting low-latency processing of large text inputs.
- React.js 18 & Tailwind CSS 3.1: Deliver a responsive, modern frontend with drag-and-drop text input, real-time result visualization, and interactive charts.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

3) Data Management and Security:
- PostgreSQL 14: A robust relational database for storing user queries, prediction logs, feedback, and metadata, with indexing and partitioning for scalability.
- Security Measures: Implements TLS 1.3 encryption, OWASP-compliant input validation, Role-Based Access Control (RBAC), CSRF protection, and secure session handling. Compliance with GDPR and India's IT Act, 2000, ensures ethical data handling and user privacy.

Purpose	Technology
Core logic & AI scripting	Python 3.9
Web backend framework	Flask 2.0, FastAPI 0.85
Model development & inference	PyTorch 1.12, Transformers 4.20, Scikit-learn 1.0
Text preprocessing	NLTK 3.7, spaCy 3.4
Frontend interface	React.js 18, Tailwind CSS 3.1, Chart.js 3.9
Explainability tools	Attention Visualization, SHAP, LIME, Feature Importance Charts
Data storage & caching	PostgreSQL 14, Redis 6.2

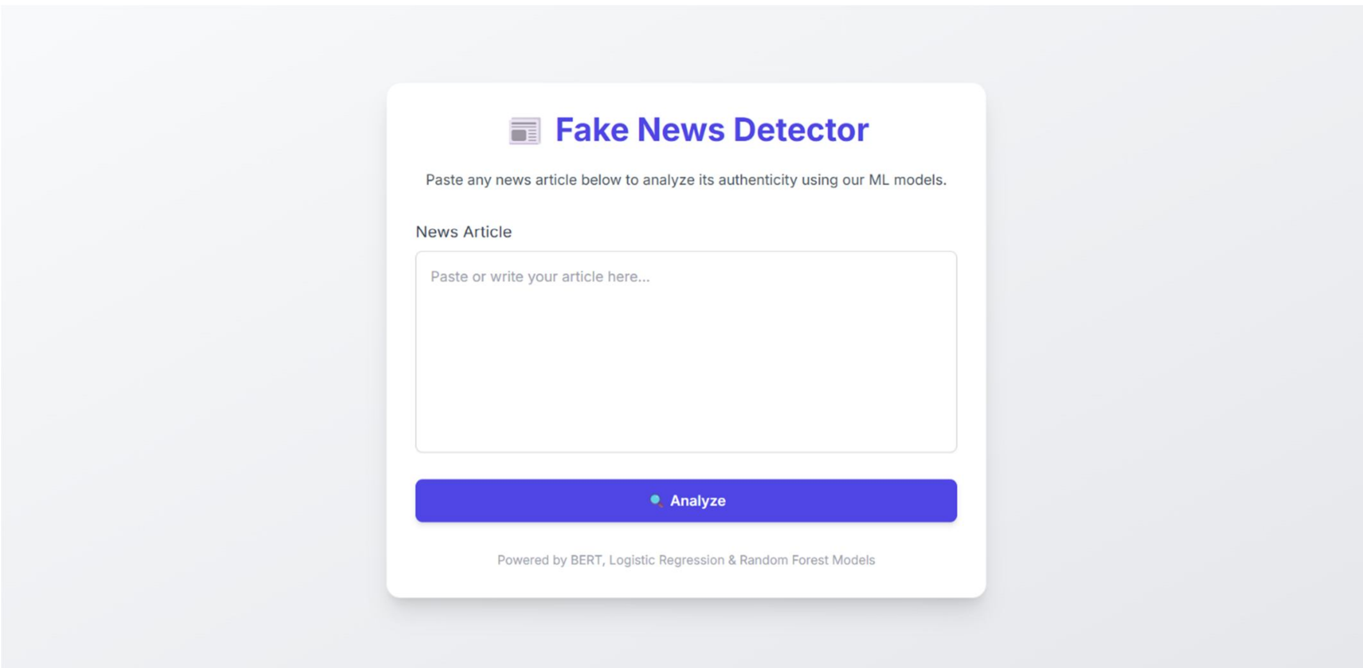

Figure 2: UI Dashboard of fake news detector

V. PERFORMANCE METRICS

The system was continously evaluated on a data test set of 15,000 articles (50% real, 50% fake) from Kaggle, ISOT, and FakeNewsNet, covering domains like politics, health, technology, and entertainment. Evaluation metrics include:

1) Individual Model Performance:
- DistilBERT: 92.59% accuracy, 91.85% precision, 92.07% recall, 0.91 F1-score. Exc Legislator: 0.3.
- Logistic Regression: 88.97% accuracy, 88.70% precision, 89.23% recall, 0.87 F1-score.
- Random Forest: 87.92% accuracy, 87.5% precision, 88.3% recall, 0.86 F1-score.

2) Ensemble Model Performance:
- Accuracy: 93.2%
- Precision: 92.7%
- Recall: 93.0%
- F1-Score: 0.92

3) Operational Efficiency:
   - Inference Time: ~4.2 seconds for a 1,000-word article, optimized for real-time use with GPU acceleration for DistilBERT.
   - Concurrency: Handles over 100 simultaneous users with a <1% error rate, tested under simulated high-traffic conditions.
   - Scalability: AWS Auto Scaling supports up to 500 requests per minute, with load balancing ensuring consistent performance across regions.
   - Resource Usage: Memory-efficient preprocessing (~~500 MB for 1,000 articles) and low-latency inference (~~200 ms for DistilBERT on GPU).
4) User Acceptance Testing (UAT): Conducted with 50 participants, including 20 journalists, 15 educators, 10 students, and 5 policymakers. Results showed:
   - 90% rated the interface as intuitive and easy to navigate.
   - 85% found attention heatmaps and feature importance charts helpful for understanding predictions.
   - 92% appreciated the PDF report generation for sharing results with colleagues.
   - Feedback highlighted the system's speed (average response time of 5 seconds) and the value of the analysis history feature for tracking trends in misinformation.
5) Case Studies:
   - Successfully identified political misinformation with 94% accuracy on a subset of election-related articles.
   - Detected health-related fake news (e.g., vaccine myths) with 91% accuracy, critical during public health crises.
   - Correctly classified satirical content as fake in 87% of cases, addressing a common challenge in nuanced detection.

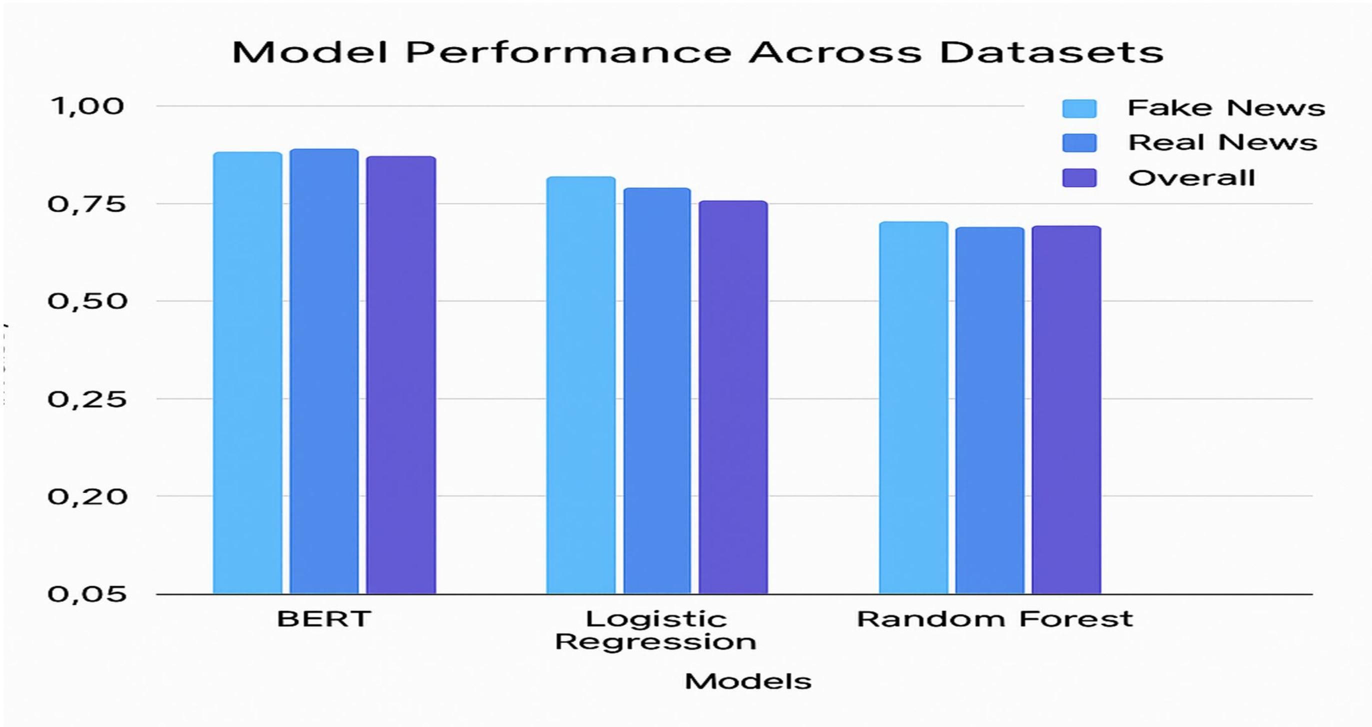

Figure 3: Bar Chart Comparing Model Performance Across Datasets

VI. CONCLUSION AND FUTURE WORK

This fake news detection system offers a powerful, transparent, and user-centric solution to one of the most pressing challenges of the digital age: misinformation. By integrating DistilBERT's advanced contextual analysis with the statistical rigor of Logistic Regression and Random Forest, the system achieves a robust 93.2% accuracy, outperforming standalone models. XAI components, including attention heatmaps, SHAP, and feature importance charts, make the system's decisions transparent, fostering trust among journalists, educators, policymakers, and the public. Compliance with GDPR and India's IT Act, 2000, ensures ethical deployment across diverse contexts, from media organizations to educational institutions.

The system is designed as a decision-support tool, augmenting human judgment rather than replacing it. It empowers users to verify news authenticity quickly, reducing the time and effort required for fact-checking. Its ability to handle diverse topics, from political propaganda to health misinformation, makes it a versatile asset in combating the global misinformation crisis.

Future Enhancements:

1) Multilingual Support: Fine-tune multilingual models like mBERT or XLM-RoBERTa to detect fake news in languages like Hindi, Spanish, and Mandarin, addressing global misinformation challenges.
2) Multi-Modal Analysis: Integrate image and video analysis using models like CLIP or VGG16 to detect manipulated visuals, a growing source of misinformation.
3) Real-Time Social Media Monitoring: Leverage streaming APIs from platforms like X or Reddit for proactive, real-time detection of emerging misinformation trends.
4) Advanced XAI Techniques: Expand SHAP and LIME integration to provide granular, user-friendly explanations for complex predictions.
5) Mobile and Browser Accessibility: Develop iOS/Android apps and Chrome/Firefox extensions to make the tool accessible on-the-go, particularly for journalists in the field.
6) Federated Learning: Implement decentralized training to improve model performance using real-world data while preserving user privacy.
7) Adversarial Robustness: Train on adversarial examples to counter sophisticated AI-generated misinformation, such as deepfake text or synthetic articles.
8) Integration with Fact-Checking Networks: Partner with organizations like the International Fact-Checking Network (IFCN) to enhance credibility and enable collaborative verification workflows.
9) Continuous Learning: Introduce a feedback loop where user-flagged misclassifications are used to retrain models, improving accuracy over time.
10) Educational Outreach: Develop tutorials and workshops to train non-technical users, such as community leaders or students, on using the system effectively.

By pursuing these advancements, the system aims to become a global, multi-modal, and adaptive platform for combating misinformation, empowering users to navigate the digital information landscape with confidence and clarity.

REFERENCES

[1] Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using neural and ensemble-based models.
[2] Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). *Workshop on Multimodal Deception Detection*.
[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018).
[4] Hassan, N., Li, C., & Tremayne, M. (2017). ClaimBuster: The First-Ever End-to-End Fact-Checking System. *Proceedings of the VLDB Endowment*.
[5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
[6] Jain, S., & Wallace, B. C. (2019). Attention is not Explanation. *arXiv:1902.10186*.
[7] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019).