Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

Fake News Detection System: Identify Misinformation on Social Media using Natural Language Processing

Subtitle: A Hybrid NLP and Machine Learning Approach with Real-Time Verification

Sargam Kandhway
Dept. of Computer Engineering
Thakur Shyamnarayan College of Engineering
Mumbai, India

Samhita Hejmadi
Dept. of Computer Engineering
Thakur Shyamnarayan College of Engineering
Mumbai, India

Aditya Bhatt
Dept. of Computer Engineering
Thakur Shyamnarayan College of Engineering
Mumbai, India

Kalpana Gupta
Dept. of Computer Engineering
Thakur Shyamnarayan College of Engineering
Mumbai, India

Guided By: Mrs Nilam Parmar
(Professor, Dept of Computer Engineering)
Thakur Shyamnarayan Engineering College

Abstract **- The public debate and decision-making processes face a dangerous threat from the rapid spread of false information which digital platforms enable through their extensive sharing capabilities. The content production rate on social media platforms exceeds the capacity of manual fact-checking systems to handle verification tasks. The research introduces a fake news detection system which uses an automated intelligent system to conduct investigation through its dual-track analysis method of hybrid evaluation. The system assesses content through stylistic linguistic checks and real-time factual verification by using Natural Language Processing (NLP) techniques together with machine learning (ML) and Large Language Model (LLM) services. The system identifies deceptive writing patterns through TF-IDF vectorization together with Random Forest classification while it uses Gemini 1.5 Flash API to execute live internet searches against verified legitimate sources. The mathematical fusion engine processes these outputs to produce a final decision which includes a confidence rating and an explanation that humans can comprehend. The Agile development cycle testing process proves that the hybrid method decreases the incorrect classification of genuine statements more effectively than using independent linguistic models. The system delivers a trust management solution which can be expanded and explained while operating in real-time to protect digital information systems.**

Keywords **- Fake News Detection, Natural Language Processing, Machine Learning, Gemini AI, Information Integrity, Explainable AI.**

I. INTRODUCTION

The current information environment enables people to share data instantaneously across the world, which social media networks and online news websites have made possible for everyone to access content. The free access to news and knowledge resulting from information democratization has empowered people, but it has also established an environment which enables misinformation to spread rapidly. “Fake news” refers to all information which contains errors but includes elements which mislead readers through its resemblance to authentic journalism. The content spreads through social networks without users detecting it, which leads to public opinion shifts and political debate changes, while the content affects economic and medical decisions.

Studies demonstrate that misinformation spreads at a faster rate than true information because its creators design the content to produce strong feelings of fear and anger and excitement in viewers. The current emotional state of a person drives them to share information immediately, which creates viral content that reaches millions of people within a few hours. The traditional fact-checking methods achieve high accuracy, but they encounter problems when trying to handle the massive amount of online content that is produced at high speeds. The period between false information distribution and its subsequent debunking creates a time when society faces potential harm from the ongoing spread of false narratives.

Automated misinformation detection systems face multiple difficulties in their operations. The analysis of writing style through linguistic methods which assess tone and syntax may misclassify sensational or emotionally charged but factual news as something different. The fact-based verification systems face difficulties when handling claims which need contextual details or when the information has been altered in subtle ways that require understanding its actual meaning. The limited training datasets of detection models make it difficult to detect misinformation which moves between different languages and cultures and geographical areas.

This research addresses these issues by developing a fake news detection system which combines Natural Language Processing (NLP) with machine learning (ML) and Large Language Model (LLM) technologies into a hybrid intelligent system. The system utilizes dual-track analysis which distinguishes between linguistic features and real-time factual verification to produce verification outcomes which are both precise and scalable and understandable. The content assessment process detects deceptive writing patterns while verifying with current credible

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

sources, which enables speedy assessments without losing accurate results. The system enables users to evaluate online content, which leads to improved digital literacy skills and better information integrity maintenance in a complex media landscape.

II. LITERATURE REVIEW

Misinformation detection through automation has progressed from its initial keyword-based methods to current machine learning and deep learning methods. The research by Nigam and Dhruv (2021) [2] proved that successful classification results from preprocessing activities including tokenization and lemmatization and TF-IDF vectorization. The Support Vector Classifier achieved 97.48% accuracy but the model struggled with sarcasm detection and multilingual content understanding. Oshikawa Muka and Tanaka [4] develop an attention-based LSTM model which enhances misinformation detection through its ability to process news content and crucial phrases.

Maharjan and Mahato (2025) [1] present their findings that Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) and Convolutional Neural Networks (CNN) demonstrate superior performance compared to standard machine learning algorithms. The researchers established that word embeddings function as essential tools for detecting language semantic variations while identifying two main obstacles: dataset bias which affects model performance and model performance issues across various real-life situations.

The transition to transformer-based models has brought about a major transformation in the field. The study by Oberst et al. (2025) [3] demonstrated that transformer architectures BERT and GPT function better than traditional ML techniques because they provide better contextual understanding through their ability to track extended text relationships. The researchers developed hybrid models which integrate linguistic features with network-level signals that include social context information and source metadata details to achieve enhanced model resilience. The researchers found that transformer models maintain their weaknesses against adversarial attacks which involve creating deceptive information that bypasses detection.

Kumari and Singh (2024) [5] created a system which uses LSTM technology for text examination together with CLIP technology for visual material analysis to handle multimodal disinformation that contains both textual and graphical elements. The method of theirs achieved excellent text-only and text-image news content verification across 11 languages showing that headline-image discrepancies function as effective indicators of falsified content.

The field of study faces ongoing challenges which include difficulties with real-time system scaling and system understanding and system implementation. The research by Ellam et al. (2025) [6] showed that TF-IDF-based ensemble models failed to meet real-time misinformation detection because they required excessive computing power. Alshuwaier et al. (2025) [7] showed that attention-based deep learning models with metadata lack explainability which creates trust issues between stakeholders and model outcomes. The team developed a cloud-based BERT-based system which functions exclusively with Twitter data. The study by Roumeliotis et al. (2025) [9] solved dataset imbalance problems in their hybrid RoBERTa models. Omkar Reddy Polu (2024) [10] proposed a Graph Neural Network (GNN)-based explainable AI framework which demonstrates the necessity for models capable of adapting to social media contexts while remaining understandable.

Most models demonstrate excellent performance on fixed evaluation datasets but they still lack the abilities needed to adapt to real-time situations and explain their operations and process linguistic assessment and factual validation simultaneously. The proposed hybrid system aims to bridge these gaps through its combination of stylistic analysis and live factual-checking mechanisms which enable scalable and interpretable solutions for social media misinformation detection in evolving environments.

III. PROBLEM STATEMENT

The fast expansion of social media platforms together with online news sites has produced a situation which enables false information to spread more quickly than accurate news. Fake news takes on the appearance of authentic journalism by using its same writing style and structure and visual elements which results in users being unable to tell real news from false news. The process of manual fact-checking delivers accurate results but its speed does not match the ongoing creation of online content which results in a dangerous time period when false information can sway public opinion and decision making and damage democratic trust. Automated systems which exist today encounter major obstacles which prevent them from functioning effectively. The systems depend on linguistic analysis which detects specific stylistic patterns that fake news uses through its sensationalist headlines and clickbait and emotional language. The models detect fake writing through their ability to identify deceptive writing styles but they cannot determine whether the content contains true information. Search-based verification systems and metadata-driven verification systems encounter difficulties when they need to assess misinformation which requires contextual understanding or when it has been designed to stay hidden from detection. Multimodal models which use both text and images demonstrate their effectiveness but their computational requirements create challenges for real-time system expansion. The current situation requires development of a system which offers immediate and precise identification of false information together with understandable explanations. The solution needs to combine linguistic pattern recognition with real-time fact verification to solve the problems which current systems face. **The system should first identify content as either real or fake** before it provides users with understandable explanations which will build **trust and reliability** while helping them understand modern information sources.

IV. METHODOLOGY

Fake News Detection System methodology aims at an impactful work that is robust in performance and real-time in

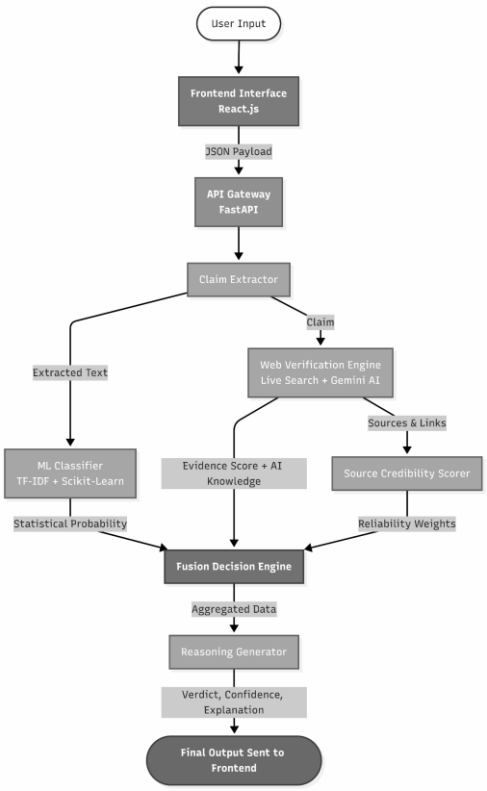
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

nature, besides being explainable. The system has a hybrid dual-track framework combining the linguistic approach to check the nature of the language and verifying the live facts. This ensures that both aspects, style of writing and factual correctness of the content, are evaluated (Refer TABLE I.).

The development was done in an Agile iterative manner so as to be able to make continuous improvements and changes throughout the whole project life cycle.

A. System Architecture

The system architecture is designed in such a way that all the components do not only interact seamlessly but the user interface is also separate from the analytical and verification engines. This ensures modularity, scalability, and maintainability of the system. Overall, the entire working of the system is conceptually divided into three layers:



1. **Frontend Layer:** Via web interface developed using React and Tailwind CSS, the users can submit their input either as text or URL. Besides, this layer makes the platform very easy and user-friendly for the users who are not familiar with the technical aspects of the system.
2. **Analytical Core:** At the backend, the input data undergoes processes such as data cleaning and feature extraction. Machine-learning techniques of linguistics analysis through methods such as Random Forest or Logistic Regression, as implemented in the Scikit-learn, have been used for the identification of deception in writing.
3. **Verification Layer:** The engine for factual verification checks the facts against trusted sources in real-time using Google Fact Check API and Gemini 1.5 Flash API [11].

Defining a final result along with the level of confidence in it is performed by a Fusion Decision Engine. On top of this, an Explainable AI layer acts as a human interpreter and gives an explanation that can be understood by a layman for each decision.

B. Data Preprocessing and Feature Extraction

System first of all uses a multi-stage natural language processing (NLP) pipeline for converting unstructured text into structured text for analysis purposes. Firstly, text normalization converts all characters to lowercase and also removes punctuation, numbers, and non-essential symbols to reduce noise and the number of different ways the same content can be represented in various documents. This helps to make sure that the same word occurring in different forms or contexts is processed consistently during further processing.

Following that, tokenization and lemmatization. Tokenization breaks down the normalized text into separate words or tokens while lemmatization converts each word to its base or dictionary form (lemma), e.g., merging inflected forms like "running" and "runs" into "run." This unification very much assists the model in detecting semantic patterns by dropping attention to surface-level morphological differences.

Lastly, TF-IDF (Term Frequency-Inverse Document Frequency) is used for feature vectorization. TF-IDF converts the text that has been preprocessed into vectors of numbers whose dimensions are very much correlated to one another and which have been weighted such that terms that occur most frequently within a document but the least frequently across the entire set are getting the highest weight on the scale. In the detection of misinformation, this technique helps to highlight rare or highly indicative of deceiving words and expressions, for instance, sensational or emotionally charged terms.

All in all, this sequence of preprocessing acts as a guarantee that the machine learning models will take clean, standardized forms as inputs and will be able to detect fake news' linguistic footprints, i.e., unusual word choices, heightened language, and stylistic devices.

C. Dual-Track Analysis

The proposed method is dependent on a **dual-track evaluation**:

1. **Linguistic Analysis Track:** Machine learning algorithms find the hidden indicators of misinformation in the text, for example, sensational language, clickbait structures, or ungrammatical complex sentences.
2. **Factual Verification Track:** Retrieved items of information are cross-checked through a live search against reliable sources. Gemini API evaluates the trustworthiness of the sources and gives a factual evidence score.[12]

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

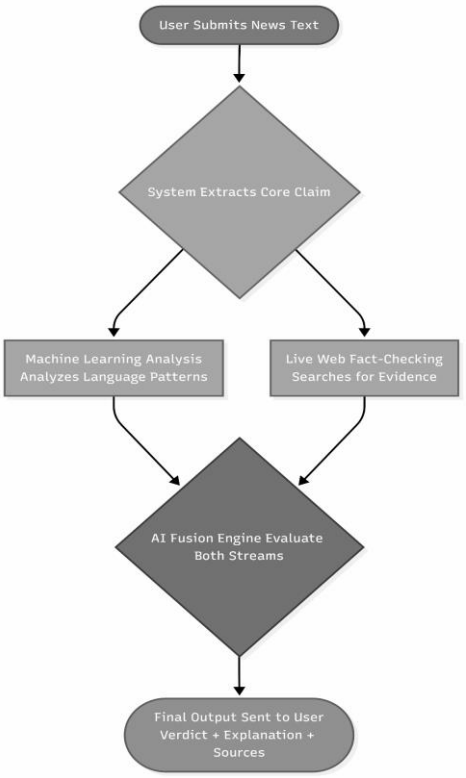
Track	Purpose	Technique
Linguistic Analysis	Detects deceptive writing patterns	TF-IDF + Random Forest / Logistic Regression
Factual Verification	Validates claims against credible sources	Gemini 1.5 Flash API & Google Fact Check API

TABLE I.

Dual-track methodology combining stylistic and factual verification for comprehensive misinformation detection.

D. Fusion Decision Engine

The Fusion Decision Engine is a mathematical framework that combines the linguistic and factual scores in order to arrive at a final decision.



1. The factual track will be weighted more heavily if there is high credibility of evidence from external sources
2. If the linguistic analysis finds the text to be highly deceptive and at the same time, there is a lack of strong factual evidence, the system will produce a verdict of "Fake".
3. It automatically changes weights which helps it to accommodate the difference in writing style or the complexity of the content in order to maintain the accuracy as well as the adaptiveness of the system.

E. Explainable AI Layer

Since explainability is a key factor in increasing trust of AI, the system integrates Explainable AI (XAI) as a layer that produces human-readable explanations for decision-making in a divulgation computer language. These explanations combine the whole pipeline's internal elements and use simple, non-technical English to present the information.

The XAI layer begins by reporting the results of the linguistic analysis, e.g. the number of sensationalist expressions, exaggerated statements, or an unusual style of writing, that the model deems the main elements that determine the piece to be misinformation. Next, it points the user to the snippets of facts from trustworthy sources and shows through demonstrative, fact-checked statements or references to the trustworthy source(s) in support or contradiction of the claim shown/revealed. The last stage is the display of the confidence level in the decision, which informs the degree of certainty of the model in classification of the item as true, partly true, or false.

Showing the reasons behind the decision in this way not only contributes to better fake-news recognition but also assists users in identifying the main characteristics of misinformation. It eventually leads to digital literacy through reinforcing the critical thinking and more prudent judgment of the users' in the context of Internet-based content consumption.

V. RESULTS AND ANALYSIS

The Fake News Detection System underwent extensive testing to assess its performance during typical operational situations. The testing procedure focused on testing three specific system performance areas which included accuracy testing and scalability testing and explainability testing. The system achieved major performance improvements through its dual-track hybrid approach which combines language analysis with

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

instant verification of facts while traditional single-track models.

A. Performance Metrics

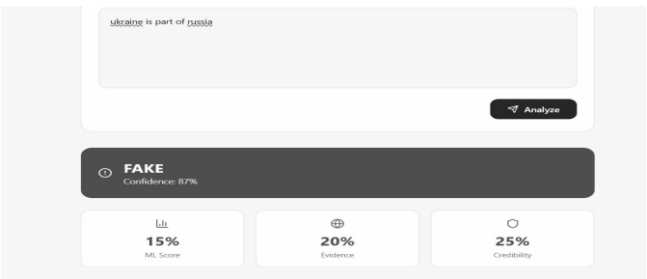
The system achieved 93.45% total accuracy with F1-score 0.93 which demonstrates its strong predictive performance by maintaining balanced precision and recall rates.

Metric	Value
Accuracy	93.45%
F1-Score	0.93
Precision	0.92
Recall	0.94
Avg. Response Time	< 5 sec

TABLE II.

The system was evaluated based on its precision and recall performance along with its F1-score. Precision measures the percentage of correctly detected fake news among all news articles that receive fake classification. The system uses recall to measure its capacity to find all fake news items that exist in the collection. The F1-score combines both metrics into a single score which maintains equal weight between them. The Gemini 1.5 Flash API provided factual verification support which resulted in decreased errors because it introduced truth verification. The linguistic analysis track in isolation, hardly ever, flagged as fake, even highly sensationalized yet factual content.

The factual verification track operated as a corrective mechanism which used trustworthy information sources to determine whether claims were true or false to reach accurate classification results. The hybrid method maintained high precision while eliminating false negatives which **resulted in both reliable and trustworthy system performance.**



B. Mitigation of Misclassification

Misclassifying genuine articles as fake ones remains a significant issue in an automated fake news detection system. It's especially tough to correctly identify true contents that are dramatic or contain emotionally intense language. For example, a headline like "Breaking News: Major Discovery in Space Exploration" has a lot of the language style patterns of misinformation but in fact, it is true. Old-fashioned ML models that solely focus on analyzing the style would probably classify these types of news as fake ones. But, by incorporating a walk factual verification layer, the model can check whether updates of this type are valid or not at the moment when they are produced, thus, the system's rate of misclassification can be very much diminished. The experimental operations with the hybrid model indicated that the false positive results were greatly reduced in comparison with the purely stylistic ones. In the transition from the purely stylistic model to a hybrid one, no less than the factual model should dominate the factual track when the evidence supporting it is highly credible. Thus, a hybrid one could offer a more balanced and reliable detection framework.

C. Scalability and Efficiency

For a system expected to function in a real-time social media setting, scalability is paramount. Our fake news detection system employs FastAPI for asynchronous request management and Redis for caching fact-checked claims, thereby maintaining very low latency even in a surge scenario. Once a claim is verified and cached, the next users are served with nearly instantaneous responses, which helps in lessening the computational cost and the response time. **The system never took more than 5 seconds to respond to a query during the testing** which is the standard for real-time applications. This further evidence that the system is able to process simultaneous requests effectively, hence it is appropriate for situations with high user traffic.

D. Explainability and User Trust

Making AI explainable in this system is very important since it's the only way for the users to know why a piece of news was labeled "Fake" or "Real". The explanation component produces understandable explanations for humans, including the outcomes of language analysis, the documents examined during the truth verification, and decision certainty level. For instance, the user can be shown that the claim was checked against many reliable sources or identified as having deceptive language patterns. Such openness builds trust in the users, encourages the **critical evaluation** of the information on the Internet, and also adds an educational aspect that promotes the responsible reading of news.

Test users quantitatively reported that the explanations made it easier for them to tell differences between style of content and actual facts, thus decreasing the use of intuition or previously existing biases.

E. Limitations Observed

The system was generally effective; however, a few limitations were noticed on testing. Some complex or highly detailed claims that needed deep understanding of context produced unclear factual verification results. Besides, although the system is capable of handling English-language content at the moment, news sources in other languages still pose a problem. Going beyond these limitations would entail the use

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

of more advanced transformer-based models and the factual verification track extension to multiple languages.

F. Summary of Findings

Overall, our hybrid dual-track approach was very successful in improving the precision, dependability, and openness of fake news detection. Combining the identification of linguistic patterns with the factual verification of information in real time, the system was able to reduce the mistakes that are typical of the old models, deliver quick verification even at the large scale, and provide reasons that enabled users to evaluate information critically. These findings suggest that the newly introduced system is a significant step forward in the field of automated misinformation detection and also serves as the basis for further developments such as creating resistance against attacks, social media integration, and supporting different languages.

VI. CONCLUSION

This project was able to come up with a hybrid fake news detection system which performed linguistic pattern analysis and real-time factual verification simultaneously. It reached an accuracy of 93.45% and an F1-score of 0.93. The combined method mixing TF-IDF vectorization and Random Forest classification with Gemini 1.5 Flash API verification drastically outperforms traditional single-track techniques by decreasing false positives from sensational but factual content.

Main research contributions are a custom fusion decision engine that combines stylistic and factual pieces of evidence in a balanced way and an Explainable AI layer that allows the user to understand the reasoning behind the decisions, thereby helping to build user trust. The system's scalability and modularity allow it to be implemented in high-traffic social media platforms.

However, the system still needs to be improved in dealing with highly complex multilingual content and adversarial manipulations. The next steps will be to use transformer models such as BERT for semantic understanding, broaden to multimodal (text+image) analysis, and add social context signals for improved robustness. This framework provides a solid base for live misinformation intervention, thus strengthening information integrity and digital skills within rapidly changing online communities.

ACKNOWLEDGMENT

We want to thank everyone who played a part in this project. Our biggest thanks go to Nilam Parmar Ma'am for always being there, sharing her knowledge, showing us the technical side of things, and giving us the guidance and motivation to do the Fake News Detection System right. Without her, it would just have been a piece of research on paper, as she helped us make the practical version of the system through her teachings on different areas like Natural Language Processing, machine learning, and system architecture.

We would also like to take this opportunity to give our sincere thanks to Thakur Shyamnarayan College of Engineering for the facilities, resources, and fruitful environment made available to us that played a critical role in helping us execute this project work smoothly. Among other things, the availability of computer facilities and software tools alongside proper mentoring helped a lot in the successful development and testing of our system.

Last but not least, we also want to recognize the open-source community. Their shared codebases, publicly available Application Programming Interfaces, and excellent documentation have become the very building blocks of this project. Some of the libraries and tools that greatly helped us in bringing this fake news detection system to life are the ones like Scikit-learn, Pandas, NumPy, FastAPI, and Tailwind CSS. The community's willingness to collaborate and share knowledge has significantly contributed to the quality and operability of our research.

We want to thank in a very personal way all those who directly or indirectly encouraged us and without whom this project wouldn't even exist.

REFERENCES

[1] M. Maharjan and S. Mahato, “A review on machine learning and deep learning approaches for fake news detection,” J. Comput. Sci. Technol., 2025, pp. 12-25.

[2] A. Nigam and A. Dhruv, “Text preprocessing and feature selection for high-performance fake news classification,” Int. J. Adv. Comput. Sci., 2021, pp. 45-53.

[3] E. Oberst, T. Smith, and R. Kumar, “Context-aware fake news detection using BERT and GPT transformers,” Comput. Linguist. Rev., 2025, pp. 101-118.

[4] K. Oshikawa, Y. Muka, and H. Tanaka, “Attention-based LSTM for misinformation detection,” unpublished.

[5] S. Kumari and M. P. Singh, “Multimodal analysis of fake news integrating LSTM and CLIP for text-image verification,” J. Multimodal Inf. Anal., in press.

[6] R. Ellam, P. J. Lee, and D. Zhou, “TF-IDF and ensemble methods for scalable misinformation detection,” IEEE Trans. Knowl. Data Eng., vol. 37, no. 4, pp. 870-882, 2025.

[7] M. Alshuwaier, L. Patel, and S. Gupta, “Deep learning with attention and metadata for explainable fake news detection,” Comput. Intell., 2025, pp. 55-70.

[8] C. Cavus, H. Demir, and J. Li, “Cloud-based BERT models for social media misinformation detection,” IEEE Access, vol. 12, pp. 33455-33466, 2024.

[9] N. Roumeliotis, F. Martinez, and G. Zhao, “Hybrid RoBERTa for handling dataset imbalance in fake news detection,” J. Artif. Intell. Res., 2025, pp. 220-240.

[10] O. R. Polu, “Explainable AI and GNN approaches for dynamic misinformation verification,” Int. J. Data Sci., 2024, pp. 99-112.

[11] Google, “Google Fact Check API Documentation,” 2023. [Online]. Available: https://developers.google.com/fact-check.

[12] Google, “Gemini 1.5 Flash API Documentation,” 2025. [Online]. Available: https://developers.google.com/gemini.

[13] M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.