VARITAS A Chatbot for Fake news detection using Machine learning and NLP

Chandani Sevkani[*], Arshi Khurshid[*], Jyoti Chaurasiya[*], Taruna Khemwani[#]

[*]Department of Computer Science Engineering, Govt. Mahila Engineering College, Ajmer, India
[#]Assistant Professor,Department of Computer Engineering, Govt. Mahila Engineering College, Ajmer, India

(22CSE026@gweca.ac.in, 22CSE011@gweca.ac.in, 22CSE052@gweca.ac.in, tarunakhemwani@gweca.ac.in)

Abstract- The rapid spread of false information through social media and communication tools is the main cause of fake news, a significant issue in today's online world. Because they rely on static data, the conventional techniques for spotting fake news are insufficient for confirming fresh information. In order to detect fake news in real-time, a novel approach based on a combination of machine learning, natural language processing, and open source intelligence techniques is presented in this study[3].

The project suggests creating a new approach to providing credibility to web-based sources of information. The new approach consists of four distinct components including an automated node analysis system that will facilitate gathering verified web-based sources of information at any given time; a fact-checking mechanism that utilizes open source intelligence through an automated search of the Internet; and a reading comprehension system that uses Natural Language Processing (NLP) techniques such as TF-IDF vectorization and cosine similarity to evaluate whether an item has sufficient context for determining its value as a source of information. The system will also provide a fallback mechanism that utilizes logistic regression to classify data regardless of the language used in the queries.

To promote transparency and accountability, all decision-making processes are logged for audit purposes, and the user interface provides an interactive experience with visualized confidence scores to enhance user trust. The hybrid type allows for greater accuracy and improved robustness when compared to conventional independent systems; this was shown through experimentation to be true for real-time falsifications of information and established cases of counterfeit news.

Keywords: Fake News Detection, Machine Learning, Natural Language Processing, Open Source Intelligence, TF-IDF, Cosine Similarity, Real-Time Fact Checking

1. Introduction

In recent years, there has been a rapid growth in digital communication, which has led to a significant increase in the spread of false information through the internet and social media platforms such as Facebook, LinkedIn, and WhatsApp. Traditional methods of spreading fake news, like newspapers and magazines, have been largely replaced by new forms of media, including social networks and websites that lack established validation processes. As a result, the challenge of effectively dealing with the spread of false information through these newer platforms has become a major concern for businesses and government organizations.

In the past, detecting fake news relied on using historical datasets and/or machine learning algorithms to predict outcomes related to specific events. While these traditional fake news detection algorithms can identify patterns found in the data they were trained on, they are not effective at predicting the emergence of new fake news. This is because they cannot perform real-time verification. Consequently, these algorithms focus only on the patterns they have learned from past events and do not seek external validation to confirm those patterns.

To address these issues, this study presents a hybrid fake news detection system that integrates Machine Learning (ML), Natural Language Processing (NLP), and Open Source Intelligence (OSINT). The system is designed to analyze text and verify claims by cross-checking them against real-time sources on the internet. This multi-step verification process is intended to improve the accuracy and reliability of news evaluation. The system supports real-time fact-checking by searching trusted fact-checking websites. It also uses NLP to compare the title of an article with its content, helping to identify misleading or clickbait headlines.

Moreover, the system is capable of functioning in various languages, making it easier to read and analyze news across different cultures. This is especially important since many countries have a diverse range of languages spoken by their populations. The system also includes a fallback machine learning model that enables it to continue functioning when external data is not available. The proposed model serves as a solution to the ongoing issue of identifying false information in the media.

2. Related Works

However, in recent times, the detection of fake news has become a major area of research because of the rapid spread of fake news over social media[1] and other online platforms. It has become essential to develop systems that can filter out fake news and detect them in order to have better public opinion formation. Various techniques have been used to make the detection of fake news more efficient. In the initial phase of fake news detection, traditional machine learning techniques like Naive Bayes, Support Vector Machine (SVM), and logistic regression were used[2]. These techniques normally consider the following aspects of the given news article and classify them accordingly:

- The frequency of the words used in the document

- The sequence of the words used in the document

- The TF-IDF score of the document

These techniques based on traditional machine learning are normally efficient, but in some cases, these techniques might not perform well unless the fake news detection system is not trained with a variety of data and fails to perform efficiently in the presence of other types of fake news

To get around these limits, people have started using methods based on deep learning. Recurrent Neural Networks and Long Short-Term Memory networks are good at understanding the order and meaning of text. These methods are more precise than older ones. But they require a lot of data and strong computing power, so they aren't useful for tasks that need to be done quickly.

Natural Language Processing is part of most fake news detection tools. It involves splitting text into smaller parts, removing common words, and converting text into numbers. People also use techniques like TF-IDF vectorization to measure how similar two texts are. This helps identify when a news headline doesn't match the actual content of the article, which is often the case with clickbait news.

Other people who do research have also tried to figure out how to use information and fact checking to make detection systems better. This is where they check if the information is correct by looking at trusted databases or information that has already been verified. Though this helps make the system better it is still not perfect because it does not check the information as it happens and it cannot find new fake news and rumors that are spreading fast.

Even though a lot of progress has been made in this area there are still some problems with the systems that exist now. Most of the systems that exist now use data so they cannot adjust to new information quickly[1]. Also most of the systems that exist now only work well with the language. Another problem, with most of the systems that exist now is that they cannot find clickbait headlines, where the headline's not true but the source is trusted.

To solve these issues, we suggest a method that uses machine learning, natural language processing for analysis, and live checks with open-source intelligence. This system supports multiple languages and includes several layers of validation. It is more useful and works well for finding fake news in real life. The method uses machine learning to spot patterns and natural language processing to understand the text. It also performs live checks using open-source intelligence to verify the information. The system is designed to be practical and effective.

3. Methodology

3.1. System Overview

The proposed Fake News Detection System is designed to apply multiple levels of verification to ascertain whether an article contains true or false information. The system utilizes numerous methods of verification in order to increase the accuracy, reliability and applicability of results obtained from the completed verification process.

The proposed fake news verification system includes a clear and coherent process that will guide the verification of everything

a user uploads to the system. A verification can consist of domain verification, real-time fact-checking, or content verification, classification via powerful machine learning protocol. In addition to identifying articles that have recently been identified as containing false information, the proposed fake news verification system will also identify previously identified articles containing false information.

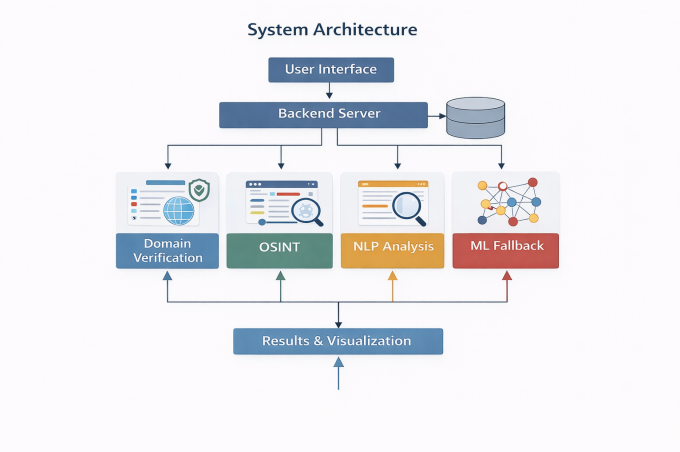

Fig. 1. System Architecture

3.2. Dataset description

This study used a dataset from Kaggle. The dataset has news statements, with labels. These labels help identify news.

The dataset includes:

(i) Text of the news

(ii)Labels that say if the news is real or fake

Here are the details of each data instance:

Statement: This is the headline or the news story. It was used for analysis.

Label: This label tells you if the statement is True or False.

3.3. System Workflow and Decision Logic

The system will have made a number of different choices (and sequentially so) for the processing of an input data (for instance, taking input a headline for a news article or a statement), or for providing a URL (as an input). Then, the data for processing that has been provided as the output will subsequently have the domain names and source of domain names checked by pre-fetching the accepted/valid domains stored in its system database. Once a valid response is made after the completion of the above-mentioned check for a domain name, an additional inspection will be made for the user-provided URL/website and processed in a similar fashion but based on an assessment made on the keywords used for finding the specific website. A further valid activity will include a machine learning method (machine learning classifier), wherein a user-provided URL/website will be classified based on previous experience of relevant and non-relevant URLs/website and provided as a response. Henceforth, the final response will have to be in sequential order, wherein multiple checks, comparisons, and standards will have been created based on the authenticity of each user URL website and processed sequentially based on the above-mentioned fashion since each and every check of the URL will have been created by the combination of different types of user URL websites and different types of data that are entered.

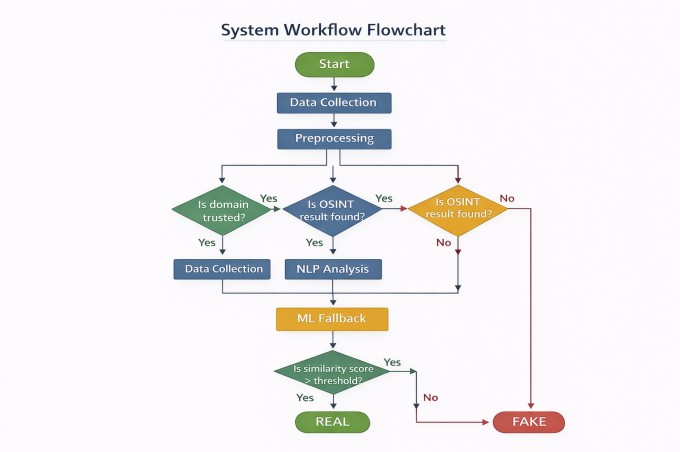

Fig. 2. Workflow Flowchart

3.4. Automated Domain Parsing and Verification

A module for parsing domains is included in the system as a way of evaluating the reliability of news sources through domain extraction from the input and comparing the extracted domain with a fixed set of trusted news sites.

As a result, verification can be performed using O(1) time complexity or constant time lookup of data. Nonetheless, just because a domain is identified as being trustworthy by the system, this does not imply that the system will rely on that attribute to determine whether the content is accurate and not misleading or manipulated. Further analysis will be conducted to determine if the content of the domains identified as trusted is correct.

3.5. Live OSINT-Based Fact-Checking

The system features real-time verification via OSINT methods in order to address the shortcomings of static datasets. Specific web searches are carried out to find accurate and reliable information from fact-checking resources, including Alt News and Boom Live.

The system uses real-time data to identify newly generated fake news and rapidly spreading false information which does not appear in its training data. The system achieves better detection accuracy through real-time data integration which enables it to maintain system performance during dynamic changes of information systems.

3.6. Web Scraping and Content Extraction

Web scraping techniques are used by the system to gather the entire article based on the URL provided by the user. As part of this process, we use HTML parsing to help find applicable text in the HTML structure of the web page.

Once the article content has been extracted, it is cleaned and preprocessed to prepare it for additional evaluation. Therefore, you will be able to evaluate the entire article rather than just the headline.

3.7. NLP-Based Text Processing and Feature Extraction

Natural language processing methods are used to analyze information presented in the form of text. In the first stage of analysis, the data undergoes initial cleaning and normalization, which entails breaking the text into individual tokens, removing common stop words, and ensuring that the dataset has a uniform appearance.

After the initial cleaning of the data, TF-IDF vectorization is used to convert the text into a numeric representation[4], which means that different texts can be compared mathematically, allowing for a more accurate analysis of the text data.

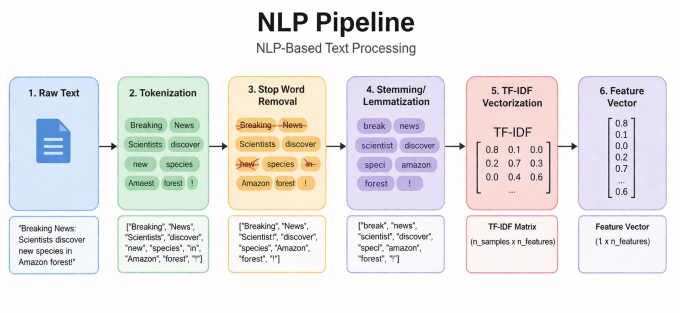

Fig. 3. NLP pipeline flowchart

3.8. Cosine Similarity for Context Validation

To determine how aligned the article is with the input text we will utilize cosine similarity. Cosine Similarity calculates the distance between 2 vectors[4] by converting the input text and article into TF-IDF vectors (a set of data representations) and computes the angle between those two vectors. The closer the angle is to zero, the more similar the texts are.

The formula for cosine similarity is as follows:

$\cos(\theta) = (\mathrm{A} \cdot B)/(\|A\|\|B\|)$.

A is the TF-IDF vector for the input text, while B is the TF-IDF vector for the article.

A high similarity score between the input text and article indicates a good match whereby it is likely that the headline accurately depicts the article, while a low similarity score may be indicative of clickbait or outright lies. By calculating the cosine similarity between the two sets of data using this approach allows us to easily identify misleading or inappropriate headlines as it relates to the actual contents of the article.

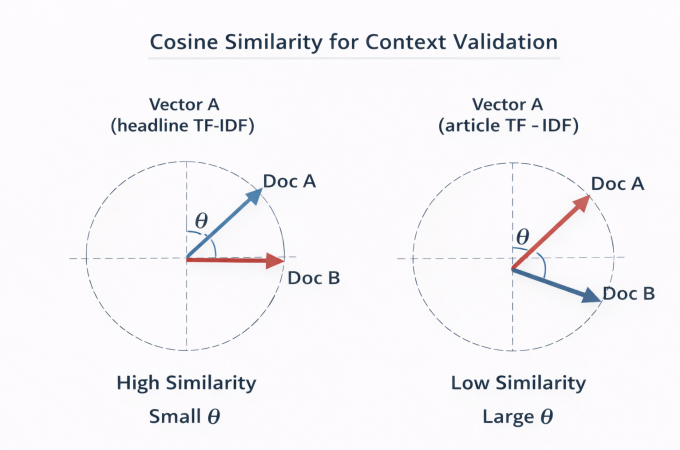

Fig. 4. Cosine similarity in 2D vectors

3.9. Keyword-Based Fallback Mechanism

Our system uses keyword matching to identify important terms in the input and weigh them in relation to known patterns of fact when input texts are too small for TF-IDF computations to yield meaningful results. This procedure will stop us from incorrectly labeling brief but legitimate entries as invalid just because there isn't enough text.

3.10. Machine Learning-Based Classification

The Logistic Regression classifier, which employs linguistic patterns and grammar analysis[4] and it is a backup technique for identifying whether news is genuine or fake.

In situations where external data sources may not be available to validate or provide alternative forms of analysis through NLP techniques, the backup component of the overall system also gives end users confidence that they will receive an accurate classification from the system when they need to rely on it.

3.11. Multilingual Support

The system enables users to type text in any language they want to use because it supports multiple languages. Thus, it will be more accessible or inclusive for multilingual-based users. The system applies natural language processing methods to analyze different words because the system can understand content that was input through any language. Information that is false or misleading gets reported between different countries and language systems, so systems need to handle multiple languages when they process this type of information. The system will gain better abilities to detect and analyze false or misleading information because it can access more chances to find such information from a larger number of people.

3.12. Audit Logging and System Transparency

The system collects information about each user's actions during the hours they are logged into the system, including the time when they logged into the system and what actions were performed after logging into the system. This data has been arranged to allow for useful continuous monitoring of the user's activities as well as to allow for a timebased evaluation of each of those user's activities.

The logging system generates several types of logs that can be used to monitor performance, identify any problems with the operations of the system, and assist with troubleshooting issues occurring with equipment used with the system. The logging system can enable assessment of decision-making processes and assist organizations in creating better decision support models and improving their operational efficiency.

The logging system will provide organizations with transparent operational activities through its operational documentation. The logging system will keep track of a user's activity and all actions taken do so will establish a historical record for future reference, which creates an organized level of accountability. The logging function creates confidence in the operations of an organization's system therefore increasing the reliability of these systems within their organizations.

3.13. Output Generation and User Interface

The output is either 'Fake' or 'Real' based on the analysis of multiple modules. The final result is made easy for the user to interpret and is well-presented. The user interface has been designed to be straightforward and easy to use; the user simply provides a piece of text or a URL to be processed and receives a result. During processing, the use of visual indicators (such as loading) helps improve the overall user experience.

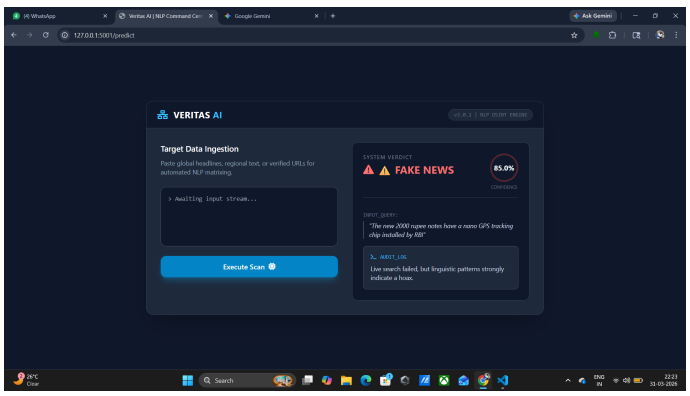

Fig. 5. Output

4. Results

In this study, an evaluation was conducted on a hybrid fake news detection system with multiple testing input sources including news headlines, short statements, and URLs. The objective was to evaluate how effectively the system identifies information as being real or fake based on a multi-layered methodology. The system was tested against both trusted news sources and misinformation that is often circulated through social media platforms such as Facebook and Twitter. Overall, results indicated that the system can process different input types of data and develop valid conclusions according to both the input data and verification method(s) applied to the data.

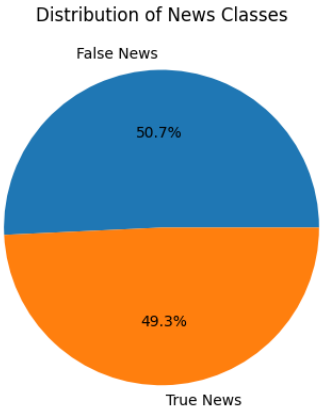

Fig. 6. Data distribution

4.1. Case-Based Analysis

After a user inputs a URL of a news article into the tool, it identifies the domain of the news source and then determines if that news source is an accepted reference for news. If the source matches an accepted news source, the tool uses Natural Language Processing (NLP) to review the content of the article to determine if it is true (1st method).

However, in cases where no news source exists, the tool has another strategy. When a user inputs a news story in text format (e.g. the text of a viral meme), without a URL or original source citation, the tool uses Open Source Intelligence (OSINT) and searches accepted news sources and tries to find definitive proof that the story is true and supports that the story is false (2nd method).

In cases where there is no original source to reference, the tool will have to rely on its own evaluation method because there is no original source for it to verify. The tool uses machine learning and NLP to analyze all available content, whether it is a website or simple text, to reach a conclusion about the reliability of a story. Thus, the tool can come up with a conclusion about whether a news story is valid or not.

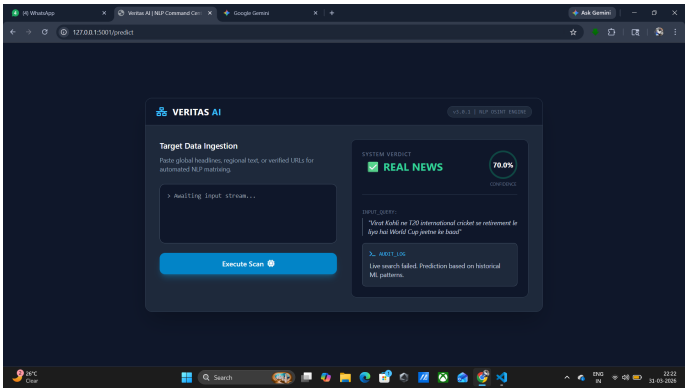

Fig. 7. User Interface

4.2. Performance Analysis

Using a hybrid approach enhances the overall reliability of the system. By using a combination of various verification methods such as verifying through domains, retrieving timely information, and using natural language processing to summarize data, this system can make an informed decision more efficiently than using a single verification method. The use of multiple methods also allows the system to handle a larger spectrum of situations and therefore perform more effectively.

The component of this system has a significant impact on the ability to perform effectively. The domain parsing module allows rapid identification of reliable information. The OSINT (open-source intelligence) module in the system allows the system to move past static datasets and immediately retrieve current data, which is extremely helpful in identifying and removing new errors, such as false claims. The NLP module improves the ability to identify misleading information by confirming discrepancies in the content of a headline and what is contained in the article. The machine learning system can also serve as a secondary confirmation system by providing another level of classification when needed.

The authenticity of an item is verified through the use of multiple verification techniques. To improve overall reliability, hybrid verification methods can be utilized in the hybrid verification system. The hybrid verification system utilizes domain verification techniques, real-time searches, and natural language processing rules-based knowledge to establish the authenticity of items.

Domain verification modules utilize the internet to allow the verification system to identify reliable sources of data quickly. At the same time, OSINT modules enable the verification process to detect the emergence of new disinformation sources. The Natural Language Processing module detects inconsistencies between the title of an item and its content; this module also provides a backup capability for items when all other methods fail.

This hybrid fake news detection model was evaluated using a test set of 2,976 samples, out of which 2,525 were correctly classified and 451 were incorrectly classified. The model achieved an overall accuracy of 84.85%, it demonstrates strong performance across both fake and real news categories.

Class	Precision	Recall	F1-Score	Support
FALSE	0.86	0.83	0.85	1510
TRUE	0.83	0.87	0.85	1466
Accuracy			0.85	2976
Macro Avg	0.85	0.85	0.85	2976
Weighted Avg	0.85	0.85	0.85	2976

Table 1. Detailed Performance Report

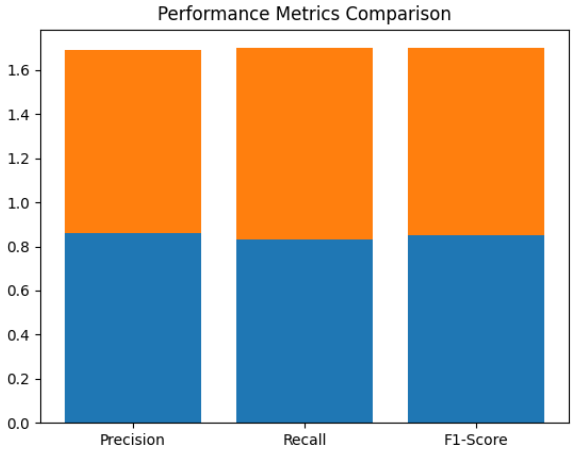

Fig. 8. Performance Analysis

4.3. Confusion Matrix Analysis

The performance of the proposed model was further evaluated using a confusion matrix, which provides a detailed breakdown of classification results.

Table 2. Confusion Matrix

	Predicted Real	**Predicted Fake**
Actual Real	1275 (TP)	191 (FN)
Actual Fake	260 (FP)	1250 (TN)

Where:

- TP (True Positive): Correctly identified real news
- TN (True Negative): Correctly identified fake news
- FP (False Positive): Fake news incorrectly classified as real
- FN (False Negative): Real news incorrectly classified as fake

This model shows the difference between real and fake news, with a balanced distribution of errors across both classes.

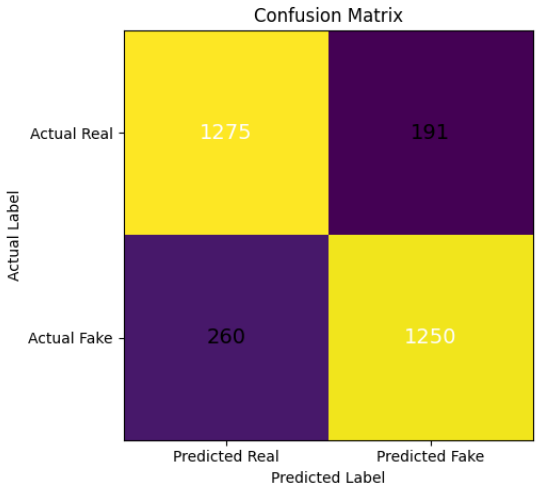

Fig. 9. Confusion matrix

4.4. Observations

The results of the testing indicated the following observations related to performance:

The system allows for efficient performance of inputs coming from verified/trusted domains by providing fast domain verification.

The OSINT modules contribute substantially toward improving detection of variants and new types of fake news articles.

Similarity analysis based on NLP will allow users to determine whether or not content is misleading/clickbait type content. The fallback (machine learning) model will allow the system to continue functioning if an external verification method is not available.

4.5. Summary of Results

Research shows that the hybrid architecture used by this new system is successful in detecting fake news because it uses many different technologies or techniques together. Because this is a "hybrid" system, it can process different types of data in many different formats or styles; This also gives the system greater flexibility, allowing it to be more efficient when used in real-world applications.

5. Discussion

Reliable output will be achieved through cooperation between different detection methods with the same accuracy level over time. The integrated use of multiple detection methods will give a unified operational capacity to produce consistent results across multiple detection methods through one means of verification; creating a unified detection system ensures the system is capable of providing dependable outputs into the future.

The decision-making architecture of the system enables high-sufficient performance using a level of operation. The domain verification module assists users in locating their trusted domains more quickly and efficiently, allowing for fewer resources and less time spent determining the trustworthiness of the information. The fact-checking mechanism consists of data obtained from publicly available sources (Open Source Intelligence) and the fact-checking system's current database of previously verified sources to support the identification of new falsehoods and viral misinformation that have not been identified using existing datasets.

A critical component in recognizing the content of news articles is natural language processing (NLP). The system uses TF-IDF vectorization and cosine similarity to measure the correspondence between the headline of an article and the body of that article to determine if they accurately represent each other. In addition, this process allows for detection of misleading information (e.g. clickbait), which might otherwise appear valid due to the established trustworthiness of the sources that created

that information. The system's keyword-based fallback mechanism provides additional support in managing shorter inputs in conjunction with providing increased overall trustworthiness.

This solution is flexible because it can be used in many different contexts, compared to current systems that rely on fixed datasets machine learning uses for detecting a fraudulent or genuine online content. The system also uses real-time OSINT verification methods that enable tracking of changes as they happen in an event being searched, so it can effectively track dynamic events. With a hybrid architecture, this solution can provide improved security from many different types of fraud detection and reduce the reliance placed on specific detection techniques.

The way that the solution is able to work effectively is based upon its ability to accept different input formats, such as multiple languages. The other means of the solution demonstrating how it works is by way of linguistic proof of its ability to work with real-world operational scenarios (i.e., false statements) in multiple languages.

Another method of determining the various steps that are used in decision-making is through the audit log. The audit log allows users of the system access to the history of decision-making, which they can then use to help with future decision-making.

The verification system that the OSINT system is based upon can only operate when there is access to the Internet; therefore, this system is limited in its operation, as it requires constant input of on-time data from sources available online to provide effective operation to its users. As a result of this, the quality and relevance of the source documents the system retrieves during its search for data will have a direct effect on the system's ability to generate accurate results; therefore, if the system retrieves results and does not provide appropriate information, the confidence in predicting based on the two (2) previous instances will reduce significantly.

There is also another limitation of the use of a fallback machine learning model that is less complex than the more advanced deep learning models. As a result, the model will promote an efficient method of operation but it may limit the ability for the Fall-back System to capture complex linguistic structures. Future enhancements could include the incorporation of advanced machine learning models into the existing Fall-back System to improve how well it performs.

This report demonstrates that the proposed hybrid system is an effective and viable method for detecting false information via fake news. The integration of various techniques, and when addressing key limitations associated with currently existing systems, the result is a more robust method of detecting fake news that can be applied in a practical way within a variety of real-life applications.

6. Conclusion

The researchers created a combined system that can tell if news is fake or not. This hybrid fake news detection (HFND) system uses machine learning techniques and natural language processing (NLP) methods to determine if something is false; it also includes open-source intelligence findings. The HFND has four separate sections that operate independently of each other. These four sections include (1) domain verification, (2) real-time fact checking (using current information), (3) NLP-based content analysis, and (4) a fallback machine learning method (that will work if all of the other methods are unable to find the truth about a specific item). All four techniques can be used in different languages, regardless of where the user is located in the world.

The results show that the newly developed method offers a high degree of accuracy of assessment when evaluating the validity of a news article, as demonstrated by the study of the ability to evaluate news article validity falsely or accurately through a multi-sourced view of both text citation sources and hyperlinks within the articles themselves. The identification of both current and novel ways of deception inside a news story has been made possible by the integration of various validation techniques and numerous independent validation methods, which has led to a significantly greater overall system reliability.

This version one is the finished product of an advanced fake news digital disinformation detection system using improved methods of analyzing text and an instant fact-checking system. This will allow for modern ways to detect and combat disinformation online, as it solves some of the problems found in prior research on detecting fake news through its comprehensive approach.

7. References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[2] Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," in Proc. 55th Annual Meeting of the Association for Computational Linguistics, pp. 422–426, 2017.

[3] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[4] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011

[5] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017.

[6] J. Thorne et al., "FEVER: A Large-Scale Dataset for Fact Extraction and Verification," Proceedings of the NAACL-HLT (North American Chapter of the Association for Computational Linguistics), 2018.

[7] S. Ruchansky, S. Seo, Y. Liu, "CSI: A Hybrid Deep Model for Fake News Detection," Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), 2017.

[8] F. A. Alshuwaier, "Fake News Detection Using Machine Learning and Deep Learning Algorithms: A Comprehensive Review," Computers, MDPI, 2025.

[9] S. Kumari et al., "A Deep Learning Multimodal Framework for Fake News Detection," Engineering, Technology & Applied Science Research (ETASR), 2024.

[10] W. Yu et al., "Fake News Detection Based on Dual Evidence Perception," Expert Systems with Applications, Elsevier, 2024.

[11] J. Jouhar et al., "Fake News Detection Using Machine Learning Algorithms," Procedia Computer Science, Elsevier, 2024.