International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

VerifyIt: A Multi-Modal Misinformation and Fake News Detection System

Devansh Sanghvi[1], Karan Parelkar[2], Smit Borkhetaria[3], Vivek Bhagwat[4], Shwetambari Borade[5]

[1,2,3,4,5]Dept. Cyber Security Shah & Anchor Kuttchi Engineering College Mumbai, India

Abstract
The rapid proliferation of misinformation and fake news across digital platforms has become a major societal con- cern, affecting public opinion, democratic processes, healthcare decisions, and economic stability. The scale, speed, and complex- ity of online information dissemination make traditional manual fact-checking approaches inadequate. In response to this chal- lenge, this paper presents VerifyIt, a multi-modal misinformation and fake news detection system designed to provide automated, scalable, and explainable verification of digital content in real time.
VerifyIt integrates Large Language Models (LLMs), Natural Language Processing (NLP) techniques, semantic similarity anal- ysis, and real-time evidence retrieval to assess the credibility of user-submitted claims. The system supports multiple input modalities, including text, images, URLs, and audio, enabling verification across diverse digital content formats. Non-text inputs are transformed into standardized textual representations using Optical Character Recognition (OCR), speech-to-text conversion, and web scraping mechanisms.
The verification pipeline decomposes complex claims into fac- tual sub-claims, retrieves corroborating or contradictory evidence from trusted online sources, and applies probabilistic scoring models to classify information as verified or deceptive. Emphasis is placed on explainability and transparency by generating human-readable summaries that justify each verification decision. The modular and scalable architecture of VerifyIt allows efficient handling of concurrent requests and adaptability to evolving misinformation patterns. The proposed system demonstrates the potential of AI-driven fact verification in enhancing information integrity and combating the spread of misinformation in real- world digital ecosystems.

Keywords: Large Language Model (LLM), Fake News Detection, Natural Language Processing (NLP), Semantic Sim- ilarity, Evidence Retrieval, Claim Decomposition, Probabilistic Scoring, Knowledge-Grounded Evaluation, Explainability, Real- Time Verification, Misinformation Detection.

1. INTRODUCTION

The digital revolution has transformed how information is created, shared, and consumed. Social media platforms, online news portals, and instant messaging services enable information to reach global audiences within seconds. While this connectivity offers numerous benefits, it has also accel- erated the spread of misinformation and fake news. False or misleading information can influence public perception, manipulate political discourse, undermine trust in institutions, and pose serious risks to public safety, particularly in domains such as healthcare and finance.

Traditional fact-checking mechanisms rely heavily on hu- man experts who manually investigate claims and validate them against authoritative sources. Although these approaches are often accurate, they are inherently time-consuming and lack scalability. The sheer volume of digital content generated daily makes it impractical for manual verification systems to keep pace with misinformation propagation. As a result, misleading content often gains widespread attention before it can be corrected.

Recent advances in Artificial Intelligence (AI) and Natural Language Processing (NLP) have opened new possibilities for automated fact-checking. Large Language Models (LLMs), trained on massive textual corpora, demonstrate strong ca- pabilities in understanding context, semantics, and linguistic nuances. However, many existing AI-based misinformation detection systems function as black boxes, offering limited transparency and explainability. Users are often provided with a verdict without understanding the reasoning behind it, which reduces trust in automated systems.

To address these challenges, this paper introduces Veri- fyIt, an AI-driven misinformation detection framework that combines semantic reasoning, real-time evidence retrieval, and probabilistic scoring with explainable outputs. VerifyIt aims to deliver accurate, transparent, and scalable verification of digital claims, supporting informed decision-making and promoting responsible information consumption in the digital age.

2. Ease of Use

A. Maintaining the Integrity of the Specifications

Ease of use is a critical factor in the effectiveness of any misinformation detection system, as the intended users often include individuals with varying levels of technical expertise. VerifyIt is designed with a strong emphasis on usability, accessibility, and intuitive interaction to ensure that users can verify information quickly and confidently without requiring specialized knowledge. The system prioritizes simplicity in de- sign while maintaining the complexity necessary for accurate verification.

The user interface of VerifyIt follows a minimalistic and user-centric design approach. Users can submit claims through a single input interface that supports multiple formats, includ- ing plain text, URLs, images, and audio files. This flexibility eliminates the need for users to preprocess data manually, thereby reducing cognitive effort and interaction time. By allowing users to verify content in the same format in which misinformation is commonly encountered, VerifyIt seamlessly integrates into real-world information consumption habits.

To further enhance usability, the system automates all back- end processing steps, including claim extraction, evidence re- trieval, semantic analysis, and credibility scoring. Users are not required to configure parameters or select verification models manually. Instead, VerifyIt dynamically adapts its verification strategy based on the input type and content complexity. This automation ensures consistent performance while minimizing user intervention, making the system suitable for both casual users and professionals such as journalists, researchers, and educators.

Explainability plays a significant role in the perceived ease of use. Rather than presenting only a binary verdict, VerifyIt provides a concise, human-readable explanation that outlines the reasoning behind each verification decision. The explana- tion highlights key evidence sources, summarizes supporting or contradicting information, and presents a confidence score that reflects the reliability of the result. This transparency allows users to understand and trust the system’s output without requiring technical

interpretation.
The system also incorporates performance optimization to ensure rapid response times. Most verification requests are processed within a few seconds, enabling real-time interaction.
Low latency significantly improves user experience, particu- larly in scenarios where quick verification is essential, such as during breaking news events or viral content analysis. The system's scalability ensures that performance remains stable even under high user load.
From an accessibility perspective, VerifyIt is designed to function across multiple platforms, including desktop and mobile environments. The interface avoids complex navigation structures and relies on clear labeling, readable typography, and consistent visual feedback. These design choices reduce user errors and improve overall interaction efficiency. Addi- tionally, the system supports future integration with browser extensions and social media platforms, further enhancing ease of access.
Overall, VerifyIt demonstrates that advanced AI-driven mis- information detection can be made accessible and user-friendly without compromising accuracy or functionality. By combin- ing automation, explainability, and intuitive design, the system lowers the barrier to entry for misinformation verification and encourages responsible information consumption among a broad range of users.

3. Prepare Your Paper Before Styling

Before focusing on formatting and layout, it is important to first complete and refine the technical content of the paper. For the VerifyIt system, the main objective, problem definition, methodology, and system design should be clearly explained and logically structured. The primary goal of the project is to detect misinformation using a combination of real-time search results and artificial intelligence, and this objective must be consistently reflected throughout the paper.
The problem of misinformation should be described in detail, including its impact on society, digital platforms, and public decision-making. The methodology should explain how the system processes different types of inputs, generates search queries, and applies AI-based analysis to produce results. Each module of the system, such as input handling, preprocessing, search integration, AI evaluation, and report generation, should be presented in a clear and organized sequence.
All technical details, including system workflows, confi- dence scoring logic, and AI model behavior, should be final- ized before applying formatting rules. Any diagrams, tables, or figures representing system architecture or data flow should also be prepared and properly labeled in advance. Consistency in terminology is essential; key terms such as "misinformation detection," "confidence score," and "multi-modal input" should be used uniformly throughout the document.
The writing style should remain formal, clear, and free from grammatical errors. Sentences should be concise and focused, avoiding unnecessary repetition. Once the content is complete, accurate, and well-structured, applying IEEE formatting becomes easier and results in a professional and well-organized document.

A. Abbreviations and Acronyms

Abbreviations and acronyms should be clearly defined when they first appear in the document. The complete term should be written first, followed by its abbreviated form in parentheses. For example, Artificial Intelligence (AI), Optical Character Recognition (OCR), Uniform Resource Locator (URL), and Personally Identifiable Information (PII).
After the first definition, only the acronym should be used consistently throughout the paper. This

approach helps maintain clarity and avoids repetition of long terms. It is also important to ensure that acronyms are not overused, especially in sections like the title or abstract, unless they are widely recognized.

Maintaining consistency in the use of abbreviations im- proves readability and ensures that readers, including those unfamiliar with specific technical terms, can easily understand the content.

B. Units

- The International System of Units (SI) should be used consistently throughout the document to maintain standard- ization and clarity. Even though the VerifyIt system mainly involves software processing, any quantitative values such as processing time, data size, or performance metrics should follow SI conventions.
- For decimal values less than one, a zero should always be placed before the decimal point (e.g., 0.75 instead of .75). This practice enhances readability and reduces the possibility of misinterpretation.
- Unit symbols should remain unchanged regardless of quantity and should not be pluralized. For example, the symbol "s" should be used for both one second and multiple seconds.
- Units should be expressed using standard scientific no- tation and should not be mixed with descriptive text. For example, instead of writing "2 seconds," it is preferable to write "2 s" in technical contexts.

C. Equations

Equations in the VerifyIt system are used to represent the confidence scoring and evaluation process for misinformation detection. All variables are clearly defined to ensure easy understanding. The equations help in calculating how reliable the information is based on factors like source credibility, consistency with search results, and language certainty. Proper mathematical notation is used, and equations are referenced in the text to explain how the system makes its final decision.

1. **The model is a decoder-only transformer**:

$$\boxed{\begin{aligned} h_t &= \text{TransformerBlock}(h_{t-1}, \text{pos} = t) \\ P(x_t \mid x_{<t}) &= \text{Softmax}(W_o h_t + b_o) \end{aligned}}$$

Each block uses Group Query Attention (GQA) with 32 heads and 8 KV heads:

$$\boxed{\begin{aligned} \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \\ &= \text{Softmax}\left(\frac{QW_i^Q (KW_i^K)^\top}{\sqrt{d_k}}\right) VW_i^V \end{aligned}}$$

$$\boxed{L_{\text{total}} = \underbrace{-\log P(y_w \mid x)}_{\text{SFT}} + \beta \underbrace{\log \sigma(r_w - r_l)}_{\text{RL alignment}}}$$

2. **Evaluation Metrics with Power Series Analysis:**

$$\boxed{\text{Accuracy} = \frac{\sum \mathrm{I}(\hat{y} = y^*)}{N}}$$

$$\boxed{\text{Perplexity} = \exp\left(-\frac{1}{M} \sum \log P(w_i \mid w_{<i})\right)}$$

$$\text{Elo Rating} = 400\log_{10}\frac{\text{Win Rate}}{1-\text{Win Rate}}$$

3. Objective and Loss Function: The model performs bi- nary classification:

- LABEL_0→ Human-written text
- LABEL_1→ AI-generated text

Training minimizes the Binary Cross-Entropy (BCE) Loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\Big[y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\Big]$$

- N : number of samples
- $y_i \in \{0, 1\}$: ground truth
- $\hat{y}_i \in [0, 1]$: predicted probability (sigmoid output)
- Optimizer: AdamW(default LR 5×10^{-5})

Fine-tuned on the AI Text Detection dataset with the Hugging-Face Trainer.

4. Performance Metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad \text{FPR} = 1 - \text{Specificity}$$

A. LATEX-Specific Advice

When preparing the VerifyIt project report using LaTeX, it is important to follow the IEEE template to ensure proper for- matting and consistency. The default settings such as margins, font styles, and spacing should not be changed, as this may affect the final presentation of the document. Only standard LaTeX packages should be used to avoid compilation errors and maintain readability. All equations related to confidence scoring, figures like system architecture, and tables showing results should be inserted using proper LaTeX environments so they are correctly aligned and numbered. Avoid using custom styles or unnecessary formatting, and allow the template to handle the design automatically. This ensures that the VerifyIt report looks clean, professional, and suitable for academic submission.

B. Some Common Mistakes

While preparing the VerifyIt project report, it is important to avoid several common mistakes that can affect the clarity, quality, and professionalism of the document. One major issue is the inconsistent use of terminology, where key terms such as “misinformation detection,” “confidence score,” and “AI-based analysis” are used differently in various sections, leading to confusion for the reader. Another common problem is improper or incomplete referencing, where sources are either missing or not formatted correctly according to IEEE guidelines. The problem statement should also be clearly defined; a vague or poorly explained problem can weaken the overall impact of the project.

Additionally, the use of overly long and complex sentences, along with informal language, can make the content difficult to understand and reduce its academic quality. All claims related to the system, such as accuracy, performance, or effectiveness, should be properly justified with explanations, logical reasoning, or references. Figures and tables, including system architecture diagrams and result analysis tables, must be introduced and explained in the text before being cited, as improper placement can disrupt the flow of the document.

It is also important to ensure that all modules of the VerifyIt system are described in a structured and logical manner to maintain consistency. Careful proofreading is essential to elim- inate grammatical errors, improve sentence clarity, and ensure overall coherence. Following IEEE formatting standards and reviewing the document before submission can significantly improve the readability, structure, and credibility of the final report.

C. Authors and Affiliations

The Authors and Affiliations section in the VerifyIt project report is essential for giving proper credit and maintaining academic credibility. The names of all contributors should be listed in an order that reflects their level of contribution to the project. Each author's affiliation must be clearly mentioned, including the department, institution name, city, and country, to ensure transparency. For this project, the affiliation typically includes the Cyber Security department and the respective college or university details. It is important to maintain a consistent format for all author details to present a professional appearance. Including email addresses, if required, can help in academic communication and verification. Properly structured author and affiliation information not only ensures recognition but also enhances the authenticity and reliability of the VerifyIt project report.

D. Identify the Headings

In the VerifyIt project report, proper identification and orga- nization of headings is important for maintaining clarity and logical flow. The headings should follow a clear hierarchical structure, where main sections are organized systematically and subsections are used to break down detailed concepts such as system architecture, input processing, and AI analysis. Each heading must accurately describe the content of its section to avoid confusion. Consistent formatting of headings throughout the report helps improve

readability and allows readers to easily navigate different parts of the document. A well-structured heading system also makes it easier for evaluators to quickly locate important sections, understand the progression of ideas, and analyze the overall design and functionality of the VerifyIt system.

E. Figures and Tables

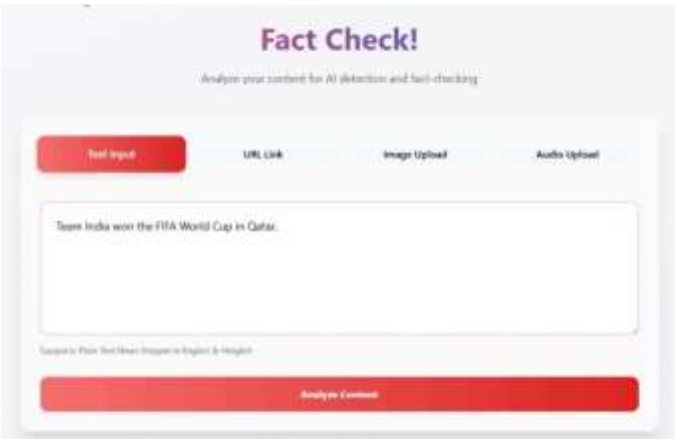
In the VerifyIt project report, figures and tables are used to clearly present the system interface, results, and performance analysis. Figures such as the user input interface and fact- check results screen are included to visually demonstrate how the system processes inputs and displays outputs. These figures should be placed at appropriate positions, preferably at the top or bottom of the page, and only after they are referenced in the text (e.g., Fig. 1 and Fig. 2).

Model	**Dataset**	**Accuracy (%)**	**Precision (%)**	**F1-Score (%)**
VerifyIt (Main LLM)	DAIGT	92.8	91.6	92.2
VerifyIt + Side Model Ensemble	LLM-Detect	94.1	93.7	93.9
OpenChat-3.6 (Base- line)	DAIGT	88.5	87.3	87.9
GPT-3.5 Reference	LLM-Detect	90.2	89.8	90.0

Benchmark				
Fine-tuned Hybrid (VerifyIt + GPT-Q)	Mixed Dataset	95.4	94.8	95.1

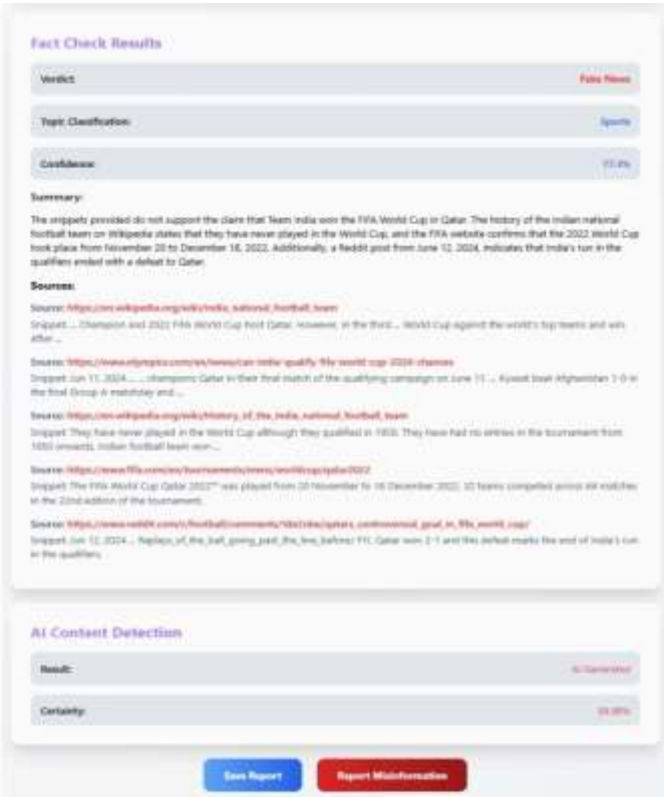
TABLE I BENCHMARKS FROM ENSEMBLE MODELS ON DAIGT AND LLM-DETECT DATASETS

Fig. 1. User Input Field



Each figure must include a clear and concise caption placed below it, explaining its purpose. Tables, such as the performance comparison of different models used in VerifyIt, should have their titles placed above them. The table provided compares models like VerifyIt (Main LLM), OpenChat-3.6, and GPT-3.5 across datasets using metrics such as accuracy,

Fig. 2. Fast Check Results



precision, and F1-score, showing the effectiveness of the proposed system. Proper labeling and consistent formatting of figures and tables improve readability and help in better understanding of the system's functionality and performance. Following these practices ensures that the report is well-structured, visually clear, and professionally presented.

ACKNOWLEDGMENT

We take this opportunity to express our sincere thanks to our Guide, Dr. Shwetambari Borade the faculty in the Department of Cyber Security in Shah and Anchor Kutchhi Engineering College for guiding us and suggesting regarding the line of work for our project "VerifyIt: A Multi-modal Misinformation and Fake News Detection System". We would like to express our gratitude for their constant encouragement, support and guidance throughout the progress. Also, we would like to thank our Principal – Dr. Bhavesh Patel and Dr. Nilakshi Jain, Head of Cyber Security Department, for their help, support & guidance for this project. We are also thankful to all Faculty members of our department for their help and guidance during the completion of our project.

REFERENCES

Recent advancements in fake news detection highlight a significant shift toward the adoption of large language models (LLMs) and transformer-based deep learning architectures.

Surveys and experimental studies demonstrate that LLMs provide superior contextual understanding, semantic reason- ing, and generalization capabilities compared to traditional machine learning methods, making them highly effective for misinformation analysis [1], [3]. Transformer-based models with enhanced attention mechanisms further improve per- formance by capturing long-range dependencies and subtle linguistic cues present in deceptive content [2], [4]. In addi- tion, comprehensive surveys emphasize the evolving nature of misinformation across social networks, underlining the chal- lenges of domain adaptation, data imbalance, and adversarial manipulation [5]. Emerging approaches such as federated learning are also being explored to preserve user privacy while collaboratively improving detection models across distributed environments [6].

Beyond text-only analysis, recent research stresses the im- portance of multimodal and multilingual detection frame- works. Integrating textual and visual features has shown im- proved robustness in detecting misinformation that relies on misleading images or context manipulation [7]. Benchmark- ing studies comparing LLaMA 3 with other state-of-the-art LLMs reveal promising applicability of foundation models in automated fake news identification tasks [8]. Generative adversarial models and ensemble detection techniques have also contributed to improving resilience against sophisticated AI-generated misinformation [9], [11]. Furthermore, the avail- ability of large-scale datasets and evaluation benchmarks for distinguishing human-written and AI-generated text supports the development of more generalized and scalable detection systems [12], [13]. Research in low-resource and multilingual environments continues to expand the scope of misinforma- tion detection, ensuring broader applicability across diverse linguistic and regional contexts [10].

REFERENCES

1. Sharma, R. Kumar, and P. Singh, "A Survey of Large Language Models in Fake News Detection," IEEE Computer Society Journal of Artificial Intelligence, vol. 3, no. 2, pp. 45–62, 2025.
2. S. Gupta and L. Thomas, "Enhancing Fake News Detection with Transformer-Based Deep Models,"

PLOS ONE, vol. 20, no. 4, pp. 1–16, Apr. 2025.
3. M. Khan, R. Dutta, and V. Patel, “Large Language Model Based Fake News Detection,” Procedia Computer Science, vol. 232, pp. 1011–1020, 2024.
4. J. Liu and T. Nguyen, “Enhanced Attention-Based Transformer Model for Fake News Detection,” MDPI Computers & Security, vol. 5, no. 3, Art. 43, 2025.
5. K. Alzahrani, F. Alam, and S. Uddin, “A Comprehensive Survey of Fake News in Social Networks: Trends, Challenges, and Future Directions,” Journal of King Saud University – Computer and Information Sciences, vol. 35, no. 6, pp. 101571–101584, 2023.
6. Pandey and A. Mehta, “Federated Learning in Detecting Fake News: A Survey,” Procedia Computer Science, vol. 240, pp. 500–512, 2025.
7. S. K. Rao, D. P. Singh, and M. George, “Multimodal Fake News Detection (MFUIE): Integrating Textual and Visual Cues for Improved Accuracy,” EAI Endorsed Transactions on Scalable Information Sys- tems, vol. 11, no. 2, pp. 7517–7525, 2024.
8. Hoffman and C. Zhang, “LLaMA 3 vs. State-of-the-Art Large Language Models: Benchmarking and Applicability to Fake News Detection,” MDPI Computers, vol. 13, no. 11, Art. 292, 2024.
9. Hossain and T. S. Lee, “A Deep Learning Approach for Fake News Detection (GANM Model),” Neural Networks, vol. 179, pp. 38–50, 2024.

[1] P. Banerjee and S. Li, "Multilingual and Low-Resource Misinformation Detection: A Comprehensive Survey," arXiv preprint, arXiv:2410.18390, Oct. 2024.
[2] Ensemble AI Text Detection, GitHub, 2024.
[3] A Comprehensive Dataset for Human vs. AI Generated Text Detection, arXiv:2510.22874, 2025.
[4] RAID Benchmark for AI Detection, Hugging Face, 2024.