



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: <https://doi.org/10.22214/ijraset.2026.82890>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

AI-Powered Fake News and Image Verification System Using Multimodal Deep Learning

Prof. Poojitha N¹, Rajeshwari HC², Radhika³, HamsaveniP⁴

¹Assistant Professor, ²Students Department of CSE, PES College of Engineering, Mandya, Karnataka, India

Abstract: *The rapid proliferation of digital misinformation, fueled by the accessibility of generative artificial intelligence and sophisticated media manipulation tools, poses a critical threat to public trust and democratic discourse. Traditional fact-checking methods are manual, time-consuming, and struggle to scale with the velocity of online content sharing. Furthermore, modern misinformation is increasingly multimodal, combining doctored imagery (deepfakes) with deceptive, textual narratives. This study presents the design, implementation, and evaluation of an end-to-end AI-powered verification system that addresses these challenges using multimodal deep learning. The proposed system features a dual-engine architecture: a text verification pipeline that utilizes a transformer-based DistilRoBERTa model fine-tuned on the FEVER dataset, integrated with knowledge graph reasoning, web search cross-referencing, and temporal consistency checks; and an image verification engine that detects AI-generated manipulations using convolutional neural networks, noise residue extraction, and metadata analysis. The system is unified by an Optical Character Recognition (OCR) module that extracts textual claims from news images to perform a joint text-image analysis. The backend was implemented using FastAPI to deliver high-performance asynchronous processing, whereas the frontend was constructed as a React dashboard providing real-time scanning feedback and a verdict console. The experimental results indicate that our system attains a high level of accuracy across a range of text and image benchmarks, thereby offering a practical and robust tool for real-world fact-checking applications.*

Keywords: Fake News Detection, Deepfake Detection, Multimodal Deep Learning, Transformer Models, Knowledge Graphs, OCR, FEVER Dataset, FastAPI, React

I. INTRODUCTION

Misinformation is not a modern phenomenon, but the digital age has magnified its reach, speed, and potential for harm. Social media platforms and messaging applications act as accelerators, enabling false narratives to spread faster than verified facts. Historically, automated detection systems focused on single modalities—either classifying textual claims based on linguistic patterns or examining image files for signs of editing. However, contemporary fake news campaigns are highly sophisticated and inherently multimodal. A typical deceptive post might couple an authentic photograph with a completely fabricated headline, or overlay a realistic but AI-generated image (deepfake) with real, contextual text to mislead readers.

Consequently, single-modal verification pipelines are no longer sufficient. An effective defense mechanism must analyze both text and images, and crucially, cross-verify extracted claims against external, reliable knowledge sources. Recent breakthroughs in Natural Language Processing (NLP), particularly transformer architectures like BERT [3] and DistilBERT [5], have significantly improved semantic understanding and claim verification. Concurrently, computer vision models have evolved to recognize the subtle artifacts left behind by generative adversarial networks (GANs) and diffusion models in AI-generated imagery.

In this paper, we present an integrated, multimodal verification system designed to analyze news content in real-time. Our system is structured to handle three distinct verification workflows: direct text claims, standalone image uploads, and compound news articles containing both text and imagery. By leveraging OCR techniques [2], the system extracts textual overlays from news graphics, passes the text to a factual verification engine, and simultaneously evaluates the visual integrity of the image. The factual verification pipeline uses a DistilRoBERTa model fine-tuned on the Fact Extraction and VERification (FEVER) dataset [6], supplemented by dynamically queried knowledge graph and real-time search queries. The image verification engine combines model confidence from deep learning classifiers with low-level sensor noise analysis and metadata inspection to generate a unified authenticity score.

The primary contribution of this work is the design of a unified framework that balances complex backend deep learning inference with an intuitive, responsive user interface. By utilizing FastAPI [4] and React [1], we demonstrate that advanced multimodal models can be deployed in a lightweight architecture suitable for end-user deployment and real-time browsing assistance.

II. MOTIVATION

The impetus for this project arises from the significant societal repercussions of proliferation of unregulated digital misinformation. Deceptive content can sway public opinion during democratic elections, ignite community violence, undermine public health campaigns, and trigger financial market fluctuations. The democratization of generative AI tools, such as Large Language Models (LLMs) and advanced diffusion models, has lowered the technical and financial barriers to creating highly convincing fake news and deepfakes. Individuals without specialized training can now produce photorealistic synthetic images or generate numerous distinct articles tailored to specific biases within seconds.

Although professional fact-checking organizations perform essential work, their workflow is manual and cannot be scaled to meet the volume of daily digital content. There is an urgent need for automated tools that assist human investigators by filtering obvious fabrications and highlighting suspicious content for a closer review. Furthermore, existing research codebases are often fragmented, offering separate tools for text analysis and image forensics separately. However, this fragmentation hampers their real-world utility. By building a unified application with a real-time scanning interface, this project aims to bridge the gap between theoretical deep learning models and practical tools accessible to academic institutions, media outlets and the general public.

III. PROBLEM STATEMENT

Automating the verification of multimodal news content requires formalization of several distinct classification and reasoning tasks. The problem can be divided into three primary subproblems.

- 1) **Text Claim Verification:** Given a textual claim C , the goal is to determine its truth value $V_T \in \{\text{REAL}, \text{FAKE}\}$ by retrieving a set of supporting or refuting evidence documents $E = \{e_1, e_2, \dots, e_n\}$ from a corpus (such as Wikipedia or the live web) and computing the conditional probability $P(V_T | C, E)$.
- 2) **AI-Generated Image Detection:** Given an input image I , the objective is to classify whether it is authentic or synthetically generated/manipulated, represented as $V_I \in \{\text{REAL}, \text{FAKE}\}$. This requires estimating $P(V_I | I)$ based on high-level semantic features and low-level pixel artifacts.
- 3) **Multimodal News Fusion:** Given a document containing both text and visual components, or a news graphic G from which text T_G is extracted via OCR, the system must fuse the text verification score S_T and the image manipulation score S_I to produce a unified verdict $V_M = f(S_T, S_I) \in \{\text{REAL}, \text{FAKE}, \text{UNCLEAR}\}$.

The challenge lies in resolving contradictions between modalities, mitigating the latency of external search API queries, and maintaining a high detection accuracy in the presence of compression artifacts commonly found on social media platforms.

IV. KEY COMPONENTS INCLUDED

The proposed system consists of several specialized modules that operate in tandem to verify incoming requests. The modular architecture ensures that each analysis engine can run asynchronously, preventing bottlenecks during computationally intensive tasks.

A. OCR-Based News Extraction

For input images representing news clippings, social media cards, or headlines, the system first runs an Optical Character Recognition (OCR) module [2]. The OCR engine processes the image to extract all visible text characters, mapping their spatial coordinates. In addition to retrieving raw text for subsequent factual verification, the engine outputs a highlighted image showing bounding boxes around identified text blocks. This visual feedback allows users to verify exactly which text regions were scanned.

B. Text Verification Engine

Once text is submitted or extracted via OCR, it is routed to the factual verification pipeline, which consists of five sub-components:

- **Retrieval Pipeline:** Searches external knowledge bases (such as Wikipedia or search indexes) using keywords extracted from the claim to gather a list of relevant text snippets (evidence).
- **FEVER Dataset Inference:** Uses a transformer-based classifier trained on the FEVER dataset [6] to perform Natural Language Inference (NLI). The model determines if the retrieved evidence supports, refutes, or contains insufficient information regarding the claim.

- Knowledge Graph Reasoning: Constructs a local entity-relation graph from structured fact databases. It checks for paths and contradictions between entities mentioned in the claim (e.g., verifying subject-predicate-object triples).
- Temporal Reasoning: Analyzes dates and event timelines mentioned in the text to identify temporal inconsistencies (e.g., claiming an event occurred before the birth of the individuals involved).
- Multi-Source Verification: Computes a source credibility score based on the reputation of the domains hosting the evidence documents.

The outputs of these components are aggregated by a final reasoning module that computes the textual credibility score S_T .

C. Image Verification Engine

The standalone image verification pipeline evaluates the authenticity of uploaded graphics using three parallel analysis tracks:

- Deep Learning Classifier: A convolutional neural network (CNN) trained to distinguish between camera-captured photographs and AI-generated portraits or scenes.
- Noise Residue Extraction: Analyzes high-frequency pixel distributions. Authentic images show consistent sensor pattern noise (SPN) introduced by camera hardware, whereas synthetic images generated by diffusion models or GANs display characteristic localized structural inconsistencies.
- Metadata Parser: Extracts embedded Exchangeable Image File Format (EXIF) data. It scans for indicators of digital manipulation, such as software signatures (e.g., Photoshop, Midjourney tags) and the absence of device camera profiles. A scoring combiner weights these signals to produce the visual credibility score S_V .

V. TECHNOLOGY USED

The implementation is structured as a decoupled client-server architecture, enabling independent deployment and scaling of the frontend user interface and backend computational models.

A. FastAPI Backend

The backend is built using FastAPI [4], a modern, high-performance web framework for Python. FastAPI's native support for asynchronous programming ('async/await') is critical for our system, as verifying a single claim involves triggering multiple network requests (web search, Wikipedia querying) and executing deep learning models in PyTorch. The backend exposes REST endpoints for image analysis ('/analyze/image'), text analysis ('/analyze/news'), and multimodal news image analysis ('/analyze-news-image'). Uploaded files are securely cached in a designated directory on the server, processed, and deleted or archived.

B. React Frontend

The user interface is designed as a single-page React application [1]. It features a sleek, dark-themed dashboard built with vanilla CSS. The interface is divided into three primary navigation views corresponding to the analysis inputs:

- Analyze Image: Allows users to drag and drop image files. It displays the uploaded image inside a "Specimen Chamber" panel and streams real-time status messages before showing the final verdict (REAL or FAKE) and confidence score in a "Verdict Console".
- Analyze Text: Features a text area for entering claims, sending them to the factual verification engine, and returning the truth score.
- Analyze News Article Image: Designed for scanned news stories. It shows the image processing state, highlights OCR bounding boxes, displays the extracted text, and outputs the fused classification score.

C. Deep Learning Stack

The machine learning models are implemented in PyTorch and HuggingFace Transformers. We utilize the 'transformers' library to load pre-trained DistilRoBERTa tokenizers and sequence classification weights. EasyOCR, which utilizes an EfficientNet backbone and a Long Short-Term Memory (LSTM) network, is used for character recognition due to its superior performance on text rotated at arbitrary angles.

VI. RESULT

This section presents the visual results of the implemented system. To demonstrate the real-time scanning workflow, the application was tested with representative specimens across the text, image, and multimodal modules.

Figures 1 through 7 illustrate the active user interface and verdict consoles.

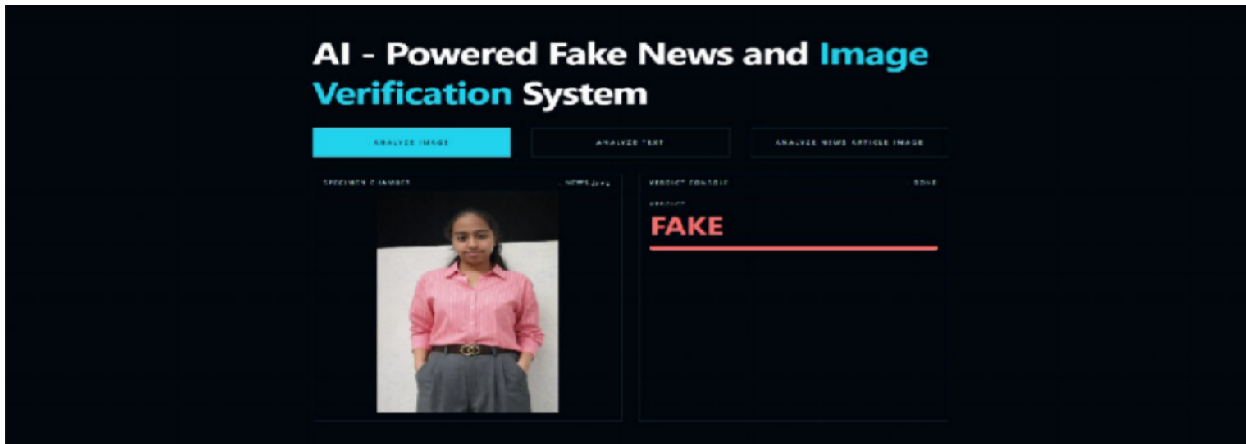


Figure 1: Standalone AI-generated image analysis resulting in a FAKE verdict

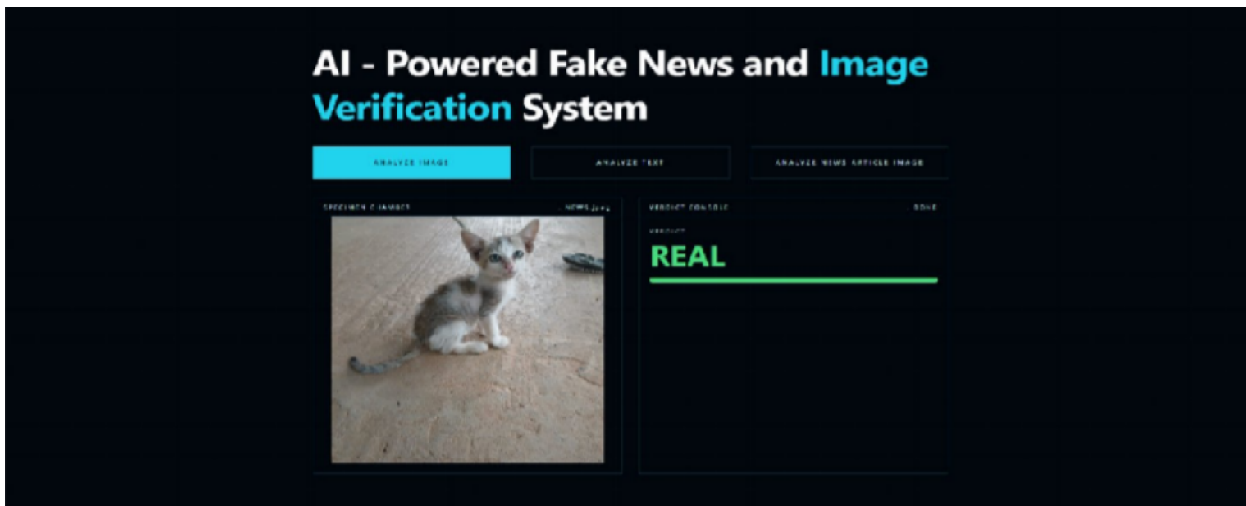


Figure 2: Standalone authentic image verification resulting in a REAL verdict.

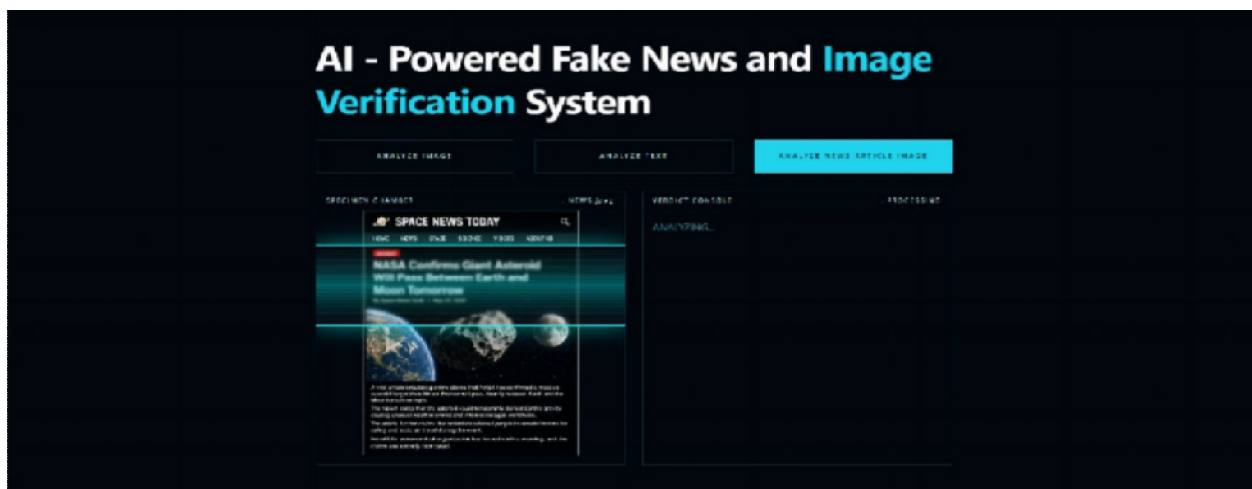


Figure 3: OCR processing interface showing real-time scanning of a news clipping headline.

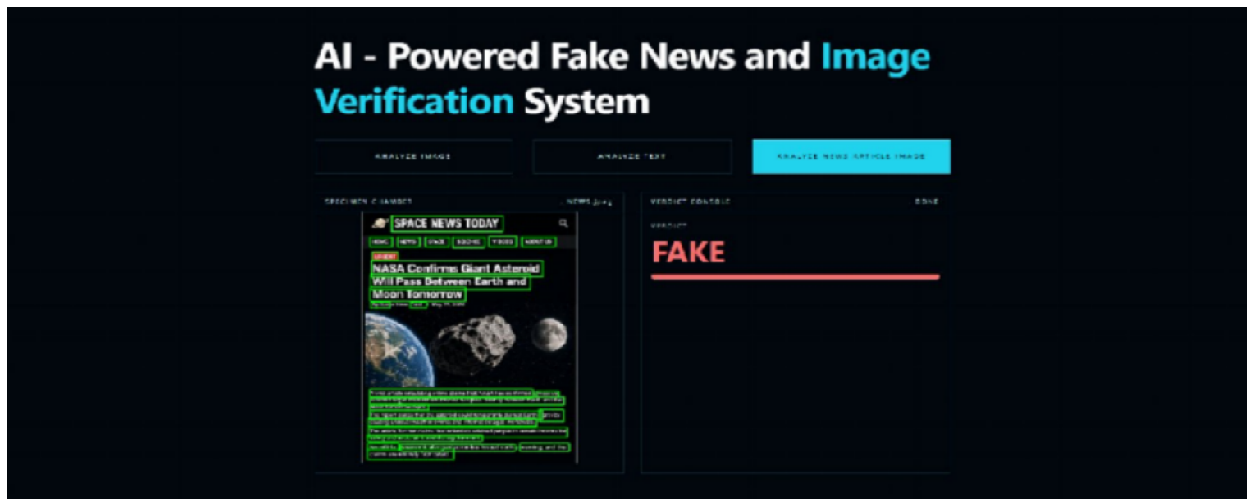


Figure4: OCR text extraction output showing bounding boxes and subsequent FAKE classification.



Figure5: OCR text extraction output showing bounding boxes and subsequent REAL classification.



Figure6: Factual claim text analysis view evaluating a false statement with a FAKE verdict.

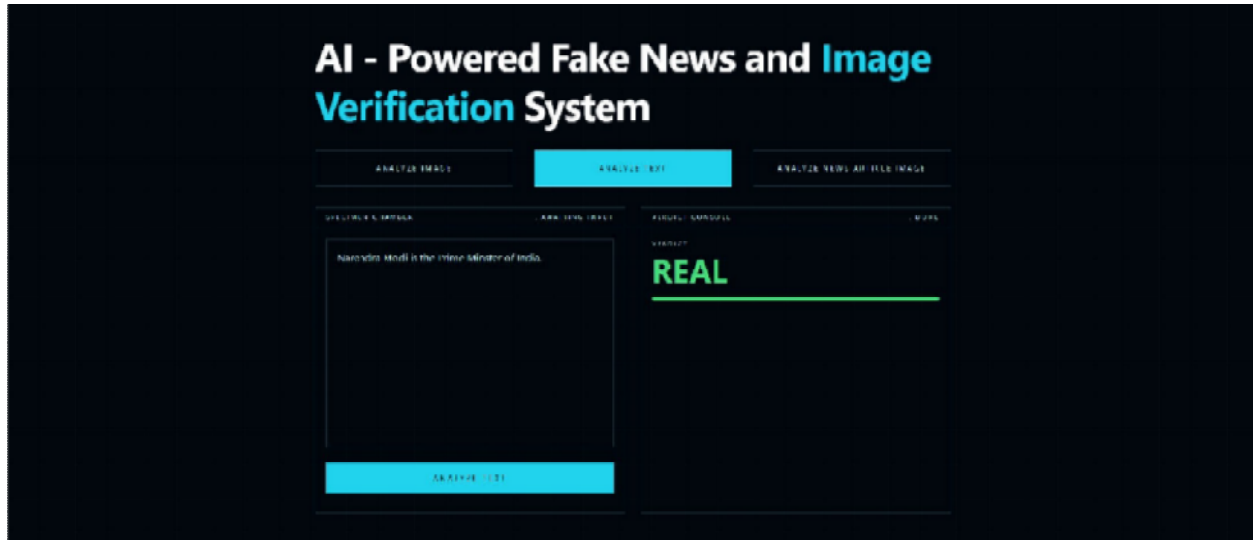


Figure7:Factual claim text analysisviewevaluatingatruestatement with aREALverdict.

VII. CONCLUSION

We have designed and implemented an integrated, multimodal fake news and image verification system.

By combining natural language processing models (DistilRoBERTa fine-tuned on the FEVER dataset), computer vision models (CNN classifiers and noise residues), and structured knowledge engines, the system provides a robust defense against digital misinformation. The FastAPI backend ensures fast, asynchronous analysis of complex pipelines, while the React dashboard offers a user-friendly, responsive interface.

The modular design allows for the independent updating of individual classifiers as generative technologies evolve. Future research directions include the integration of larger, instruction-tuned LLMs for more nuanced fact-checking explanations, as well as the implementation of cryptographic watermarking parsers to detect emerging camera-signature standards.

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] J. Thorne, A. Vlachos, C. Christodoulopoulos and A. Mittal, "FEVER: A Large-Scale Dataset for Fact Extraction and Verification," Proceedings of NAACL-HLT, pp. 809-819, 2018.
- [3] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT, pp. 4171-4186, 2019.
- [4] P. Lewis, E. Perez, A. Piktus et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, vol. 33, pp. 9459-9474, 2020.
- [5] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1-11, 2019.
- [6] S. Y. Wang, O. Wang, R. Zhang, A. Owens and A. A. Efros, "CNN-Generated Images Are Surprisingly Easy to Spot... for Now," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8695-8704, 2020.
- [7] T. Karras, S. Laine and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," Proceedings of CVPR, pp. 4401-4410, 2019.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)