IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

Multimodal AI Fake Content Detection System Using Text and Image Analysis

Mahi Gupta and Manya Vaish
Department of Information Technology
Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India
mahigupta042004@gmial.com , manya.vsh@gmail.com

Abstract: *Fake and misleading digital content has become a significant challenge in the modern information ecosystem, especially with the rapid advancement of Artificial Intelligence technologies. AI-generated text and manipulated images are increasingly being used to spread misinformation across social media and online platforms, making it difficult to distinguish between authentic and fake content. Traditional detection systems often rely on a single modality, either text or image, which limits their effectiveness in real-world scenarios.*

This paper proposes a Multimodal AI-based Fake Content Detection System that integrates both text and image analysis techniques to improve detection accuracy. The system utilizes Natural Language Processing (NLP) methods such as TF-IDF and machine learning algorithms for text classification, along with image processing techniques to identify visual inconsistencies in manipulated images. The integration of these two modalities enables a more comprehensive analysis of digital content.

The system is implemented using a Flask-based backend with a React-based user interface, providing an efficient and user-friendly platform. Experimental observations indicate that the proposed system enhances detection reliability and provides explainable outputs, including confidence scores and signal analysis.

The proposed solution aims to contribute toward reducing misinformation and improving content authenticity by leveraging multimodal Artificial Intelligence techniques.

Keywords: Fake Content Detection, Multimodal AI, NLP, Image Analysis, Machine Learning, Deepfake Detection

I. INTRODUCTION

The rapid advancement of Artificial Intelligence has significantly transformed the way digital content is created and shared across online platforms. Technologies such as natural language generation and image synthesis have enabled the creation of highly realistic text and visual content, making it increasingly difficult to distinguish between authentic and fake information. As a result, the spread of misinformation through AI-generated content has become a major concern in areas such as social media, journalism, and digital communication.

Fake content in the form of manipulated text and images can mislead users, influence public opinion, and reduce trust in digital platforms. Traditional detection methods primarily focus on a single modality, either text or image, which limits their effectiveness in real-world scenarios where content often contains both textual and visual elements. Therefore, there is a need for a more comprehensive approach that can analyse multiple data types simultaneously.

Artificial Intelligence and Machine Learning techniques have emerged as effective solutions for detecting patterns and inconsistencies in data. Natural Language Processing (NLP) methods can identify linguistic patterns in AI-generated text, while image processing techniques can detect visual anomalies in manipulated images. By combining these approaches, a more accurate and reliable detection system can be developed.

The proposed Multimodal AI Fake Content Detection System integrates both text and image analysis into a unified platform. The system is designed to provide efficient, accurate, and explainable results, making it suitable for real-

DOI: 10.48175/IJARSCT-32814

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, April 2026 Impact Factor: 8.2

world applications. The objective of this work is to improve the detection of fake content and contribute toward reducing the spread of misinformation in digital environments.

II. TECHNIQUES

Requirement Analysis

Requirement analysis is the process of identifying system needs, functionalities, and constraints required to develop an efficient and reliable software solution. For the Multimodal Fake Content Detection System, the primary requirements include accurate detection of fake content, efficient processing of both text and image data, and a user-friendly interface. The system should be capable of handling real-time inputs while maintaining high prediction accuracy and low computational cost.

Software Requirement Specification

Software Requirement Specification (SRS) defines the overall structure, functionalities, and limitations of the system. The proposed system consists of the following modules:

- Text Detection Module
- Image Detection Module
- Multimodal Decision Module
- User Interface

These components work together to analyse input data and generate reliable detection results.

A. Text Input Interface

Users can enter textual content through the interface. The system processes the text to determine whether it is real or AI-generated using NLP-based techniques.

Image Upload Interface

Users can upload images to check whether they are manipulated or fake. The system analyses visual patterns and inconsistencies.

Result Display Interface

The system displays the output in the form of “Real”, “Fake”, or “Uncertain” along with a confidence score and basic explanation.

B. Design of System

1. User Interface Layer

The frontend is developed using React, providing an interactive and user-friendly environment for input and result visualization.

2. Backend Layer

The backend is implemented using Flask, which handles API requests, processes inputs, and integrates machine learning models.

3. Processing Layer

This layer includes text preprocessing, image preprocessing, and feature extraction mechanisms required for model input.

4. Machine Learning Layer

This layer applies classification algorithms for text detection and heuristic or feature-based methods for image analysis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

C. Classification

Classification techniques are used to categorize content into predefined classes such as real or fake. For text data, machine learning algorithms such as Logistic Regression are used to classify content based on extracted features. The classification model learns patterns from training data to distinguish between genuine and AI-generated text.

D. Feature Extraction

Feature extraction plays a crucial role in improving model performance.

- Text Features: Extracted using TF-IDF, which converts textual data into numerical form based on word importance.
- Image Features: Extracted using statistical and visual analysis methods to identify irregularities in images.
- These features are then used as input for classification models.

E. Objectives

- To develop a system for detecting fake text using NLP techniques
- To analyse images for identifying manipulated or AI-generated content
- To integrate text and image analysis into a single system
- To provide accurate and explainable detection results
- To create a user-friendly interface for easy interaction

III. ARCHITECTURE

This section presents the detailed architecture of the Multimodal AI Fake Content Detection System, describing how different components interact to analyse and detect fake content. The objective of this architecture is to integrate text and image analysis techniques into a unified platform capable of identifying misleading or AI-generated content efficiently.

The architecture is designed using a modular approach, allowing independent development and scalability of each component. The system processes both textual and visual data through separate pipelines and combines the results to generate an accurate final prediction. The proposed system integrates multiple layers including user interaction, data processing, machine learning models, and decision-making components.

The proposed architecture consists of the following major layers:

A. User Interaction Layer

The User Interaction Layer acts as the interface between the user and the detection system. This layer is responsible for collecting user input and displaying the results in an understandable format.

The system provides a web-based interface developed using React, where users can easily input text or upload images for analysis.

Key functionalities of this layer include:

- User input interface for text and image
- Upload functionality for images
- Display of detection results
- Visualization of confidence scores

This layer ensures that the system remains accessible and easy to use for users with minimal technical knowledge

B. Data Collection Layer

The Data Collection Layer gathers input data provided by the user. The system supports two types of input:

- Text data entered manually by the user

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

- Image data uploaded for analysis

The collected data is forwarded to the preprocessing layer for further processing and analysis.

C. Data Preprocessing Layer

Before applying machine learning models, the input data is preprocessed to improve accuracy and consistency.

Text Preprocessing

- Removal of stopwords
- Tokenization
- Lowercasing
- Removal of special characters

Image Preprocessing

- Image resizing
- Normalization
- Noise reduction

This layer ensures that both text and image inputs are cleaned and standardized for effective processing.

D. Machine Learning Layer

The Machine Learning Layer is the core component of the system where analysis and prediction are performed.

1. Text Detection Model

The text detection module uses Natural Language Processing (NLP) techniques to analyse textual data. TF-IDF is used for feature extraction, and Logistic Regression is applied for classification. The model identifies patterns and linguistic features to determine whether the text is real or AI-generated.

Output:

Classification result (Real/Fake)
Confidence score

2. Image Detection Model

The image detection module analyses visual features to detect manipulated or fake images. It uses statistical and heuristic-based techniques to identify inconsistencies such as unnatural patterns and distortions.

Output:

Detection result (Real/Fake)
Confidence score

E. Decision Layer

The Decision Layer combines the outputs from both text and image models to generate a final prediction.
If both models indicate fake → Result is Fake
If both indicate real → Result is Real
If results differ → Output is Uncertain
This layer improves overall system reliability by integrating multimodal analysis.

F. Output Layer

The Output Layer presents the final results to the user in a clear and understandable format.
Outputs include:
Final prediction (Real / Fake / Uncertain)
Confidence score
Basic explanation of detection

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

The results are displayed through the user interface for easy interpretation.

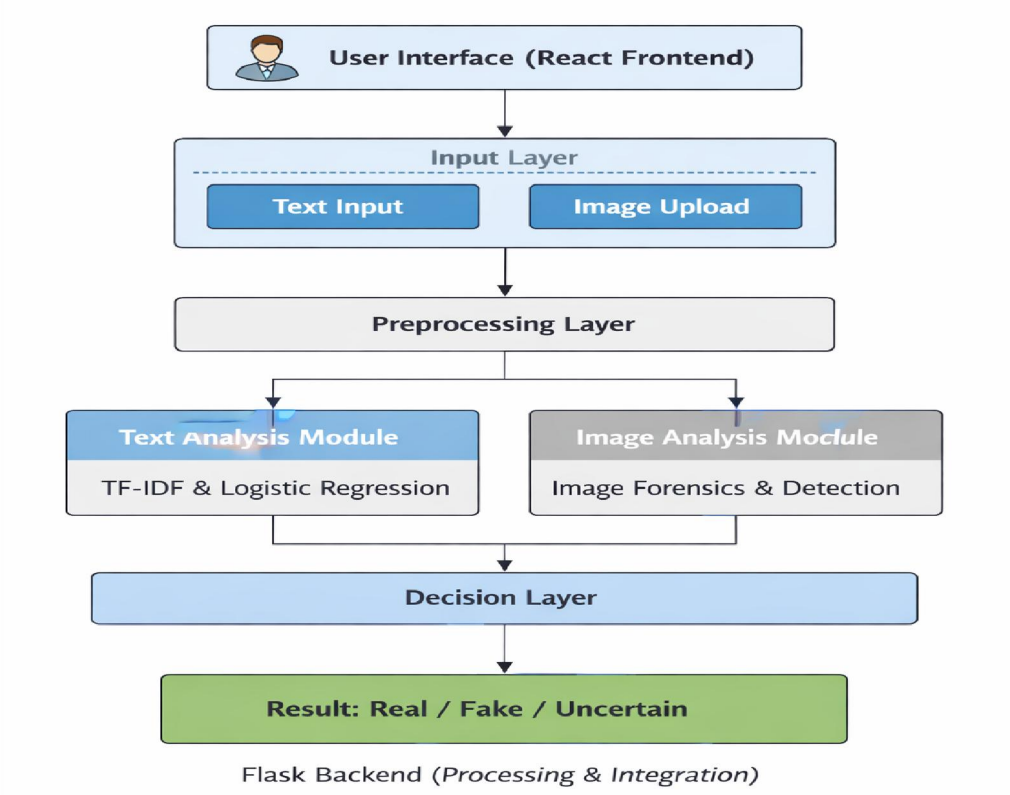

Fig 1: Architecture of Multimodal Fake Content Detection System

Architecture Flow

Step 1: User provides input (text or image)
Step 2: Data is collected and preprocessed
Step 3: Features are extracted
Step 4: Machine learning models analyse the data
Step 5: Outputs from models are combined
Step 6: Final result is displayed to the user

Advantages of Proposed Architecture

Supports both text and image analysis
Modular and scalable design
Improved detection accuracy
Efficient processing for real-time usage
User-friendly interface

IV. OVERVIEW OF THE SYSTEM

A. Input Collection

The system collects input data from the user in the form of text or images. Users can either enter textual content or upload images through the user interface. This data serves as the primary input for fake content detection.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

B. Data Processing

The collected data undergoes preprocessing to improve accuracy and consistency. Text data is cleaned using techniques such as tokenization and stopword removal, while image data is processed through resizing and normalization. This ensures that the input data is suitable for further analysis.

C. Feature Extraction

Important features are extracted from both text and image data. Text features are generated using TF-IDF, which converts text into numerical form, while image features are extracted based on visual patterns and statistical properties. These features are used as input for machine learning models.

D. Model Execution

The processed data is analysed using machine learning models. The text model classifies textual content using NLP techniques, while the image model evaluates visual data to detect manipulation or inconsistencies.

E. Decision Support

The system combines the outputs from both models to generate a final decision. Based on the analysis, the system classifies the content as real, fake, or uncertain. This integrated approach improves detection accuracy.

F. Output Generation

The final result is displayed to the user through the interface. The output includes the classification result along with a confidence score and basic explanation, making it easy for users to understand the result.

V. RESULTS OF EXPERIMENTS

The proposed Multimodal Fake Content Detection System was tested using sample datasets consisting of both real and fake content in textual and image formats. The objective of the experiments was to evaluate the performance and reliability of the system in detecting misleading or AI-generated content.

A. Objective of Experiments

To evaluate the accuracy of text-based fake content detection
To analyse the performance of image-based detection
To measure the effectiveness of combining text and image analysis
To assess overall system reliability

B. Data Preprocessing

Before testing, the input data was preprocessed to improve model performance. Text data was cleaned by removing stopwords, converting to lowercase, and applying tokenization. Image data was resized and normalized to ensure consistency in analysis.

Feature extraction techniques such as TF-IDF were applied to text data, while image features were derived using visual analysis methods.

C. Models Used

The system uses the following models:

Logistic Regression Model

Used for classifying textual content into real or fake categories based on extracted features.

Image Analysis Model

Used to detect manipulated images using statistical and heuristic-based techniques.

D. Interpretation

The experimental results indicate that the text detection model performs effectively in identifying AI-generated content based on linguistic patterns. The image detection module is capable of detecting visual inconsistencies, although its accuracy depends on image quality and complexity.

The integration of both text and image analysis improves overall detection accuracy and provides more reliable results compared to single-modality systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

VI. CRITICAL ANALYSIS

The performance of the Multimodal Fake Content Detection System depends on the quality and diversity of the dataset used for training and testing. If the dataset is limited or not well-balanced, the accuracy of the system may be affected.

The text detection module performs effectively for structured and commonly used language patterns; however, highly advanced AI-generated text may sometimes be difficult to detect. Similarly, the image detection module may face challenges in identifying highly realistic or high-resolution manipulated images.

Another limitation of the system is the computational cost associated with processing both text and image data simultaneously. The integration of multimodal data requires efficient resource management to maintain system performance.

Additionally, continuous updates and improvements in fake content generation techniques require the detection system to be regularly updated to maintain accuracy and reliability.

VII. FUTURE SCOPE

The proposed Multimodal Fake Content Detection System can be further enhanced by incorporating advanced deep learning techniques such as Convolutional Neural Networks (CNN) for image analysis and Transformer-based models for text processing. These approaches can significantly improve detection accuracy for complex and highly realistic fake content.

The system can be extended to support video-based fake content detection, which is becoming increasingly common with the rise of deepfake technologies. Integrating audio analysis can also help in detecting manipulated voice content, making the system more comprehensive.

Another potential improvement is the integration of real-time data processing and deployment as a web or mobile application, allowing users to detect fake content instantly. The use of larger and more diverse datasets can further improve model performance and reliability.

Additionally, the system can be integrated with social media platforms and browser extensions to automatically detect and flag misleading content, thereby helping to reduce the spread of misinformation on a larger scale.

VIII. CONCLUSION

The Multimodal AI Fake Content Detection System provides an effective solution for identifying fake and misleading digital content by combining text and image analysis techniques. The system utilizes Natural Language Processing (NLP) methods for analysing textual data and image processing techniques for detecting manipulated visuals, thereby improving the overall accuracy of detection.

By integrating multiple modalities, the proposed system overcomes the limitations of traditional single-method approaches and provides more reliable results. The system is designed to be efficient, user-friendly, and capable of handling real-time inputs, making it suitable for practical applications.

The experimental results demonstrate that the system performs well in detecting fake content, although its accuracy depends on the quality of input data and the models used. Continuous improvements and updates can further enhance the performance of the system.

Overall, the proposed system contributes toward reducing the spread of misinformation and improving the authenticity of digital content using advanced Artificial Intelligence techniques.

REFERENCES

[1] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.
[2] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv, 2013.
[3] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, 2011.
[4] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

[5] R. Szeliski, Computer Vision: Algorithms and Applications, Springer, 2010.
[6] D. Jurafsky and J. Martin, Speech and Language Processing, Pearson, 2019.
[7] OpenCV Library Documentation, https://opencv.org
[8] Flask Documentation, https://flask.palletsprojects.com
[9] React Documentation, https://react.dev