Deepfake-News Shield: Detecting False Articles and Manipulated Photos

A major project report submitted in partial fulfilment of the requirement
for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering

Submitted by

Madhav (221030283)
Soha Khan (221031049)
Harshit Thakur (221031013)
Divyam Saini (221030070)

Under the guidance & supervision of

Dr. Deepak Gupta (SG)

Department of Computer Science & Engineering and Information Technology
Jaypee University of Information Technology,
Waknaghat, Solan - 173234 (India)

Supervisor’s Certificate

This is to certify that the major project report entitled **‘Deepfake-News Shield: Detecting False Articles and Manipulated Photos’**, submitted in partial fulfilment of the requirement for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering**, in the Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat, is a bona fide project work carried out under my supervision during the period from July 2025 to December 2025.

I have personally supervised the research work and confirm that it meets the standards required for submission. The project work has been conducted in accordance with ethical guidelines, and the matter embodied in the report has not been submitted elsewhere for the award of any other degree or diploma.

(Supervisor Signature)
Supervisor Name: Dr. Deepak Gupta
Designation: Assistant Professor (SG)
Department: Dept. of CSE & IT

Date: 30 Sept 2025
Place:

Candidate's Declaration

We hereby declare that the work presented in this major project report entitled **'Deepfake-News Shield: Detecting False Articles and Manipulated Photos'**, submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering**, in the Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat, is an authentic record of our own work carried out during the period from July 2025 to December 2025 under the supervision of **Dr. Deepak Gupta**.

We further declare that the matter embodied in this report has not been submitted for the award of any other degree or diploma at any other university or institution.

Madhav
221030283

Soha Khan
221031049

Harshit Thakur
221031013

Divyam Shaini
221030070

This is to certify that the above statement made by the candidates is true to the best of my knowledge.

Dr. Deepak Gupta
Assistant Professor (SG)
Department: Dept. of CSE & IT

Acknowledgement

Firstly, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the project work successfully.

We are grateful and deeply indebted to our supervisor **Dr. Deepak Gupta, Assistant Professor (SG), Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Wakhnaghat**, for his invaluable guidance throughout this project. His profound knowledge and keen interest in **big data analytics, cybersecurity, machine/deep learning, and programming languages** greatly motivated us to carry out this work. His endless patience, scholarly guidance, continual encouragement, energetic supervision, constructive criticism, and valuable advice—together with his effort in reviewing and correcting our drafts at every stage—made the successful completion of this project possible.

We would also generously welcome each one of those individuals who have helped us directly or in a roundabout way in making this project a win. In this unique situation, we might want to thank the various staff individuals, both educating and non-instructing, who have developed their convenient help and facilitated our undertaking.

In closing, we wish to recognize and appreciate the enduring support and patience of our parents. Their unwavering encouragement has been a source of strength throughout this journey.

With gratitude,

Madhav (221030283)
Soha Khan (221031049)
Harshit Thakur (221031013)
Divyam Saini (221030070)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Meaning
IoT	Internet of Things
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
ViT	Vision Transformer
DFDC	Deepfake Detection Challenge
LSTM	Long Short-Term Memory

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Nowadays, misleading information represents a substantial risk because false information and deepfake images can spread faster than they can be validated. This material can shape public opinion, help spark unrest, and destroy reputations. Old methods used for detection frequently fail against more sophisticated methods.

This project successfully develops a multi-modal misinformation detection system that targets misinformation in both text and images. As it relates to textual information, we utilize Natural Language Processing (NLP) and the TF-IDF followed by a Linear Support Vector Classifier (SVC) to classify news articles as real or fake. As it relates to image-based information, we utilize a Vision Transformer (ViT) model that dimensions can detect even the smallest inconsistencies in the presence of a deepfake. To provide users with reliability, our system integrates the Google FactCheck API, which cross-checks the claims verified by vetted fact-checking organizations.

The solution is implemented as a web-based interface built in Flask and Streamlit, allowing users to test and predict text, images, or claims, while also receiving visual analytics, such as word clouds, sentiment graphs, and confidence scores. A MongoDB database is offered as a part of an extensible solution for logging queries.

In conclusion, the initiative offers a common lens in which to combine, detect, and corroborate fake news and deepfakes with a wellness-promoting and easy metric, thus providing a useful resource to enhance awareness and resilience against misinformation.

Moreover, the system promotes flexible design and ongoing learning. As new misinformation and deepfake types develop, the models can be retrained on new datasets to retain their accuracy and relevance. The modular approach makes it easy to add additional features, such as social media integration, multi-language support, or real-time alerts. This flexibility improves the utility of the system, while also ensuring it remains a source of value to journalists, educators, and the public as they develop systems for coping with changing information rights and digital literacy.

Chapter 01: INTRODUCTION

1.1 INTRODUCTION

We live in a world where information is just a click away. News spreads instantly through social media, websites, and online platforms. While this has made communication faster and easier, it has also created a serious problem — the rise of misinformation. From fake news stories designed to mislead people to deepfake images that look real but are completely fabricated, misinformation is shaping public opinion, creating confusion, and in some cases even causing harm.

The biggest challenge is that misinformation today is much harder to detect than before. Fake news articles often look like genuine reports, and deepfakes created with advanced AI tools can mimic real people with stunning accuracy. Most existing solutions deal with only one side of the problem — either fake news or deepfakes — but very few combine both. This gap makes it easier for misinformation to slip through and reach large audiences.

Our project aims to tackle this issue by building a multi-modal misinformation detection system**.** It combines machine learning and natural language processing to analyze news articles and detect fake stories, and uses deep learning models like Vision Transformers to identify deepfake images. To make the results more reliable, we also integrate the Google FactCheck API, which checks claims against trusted sources. Finally, everything is brought together in a simple web application where users can input text, upload images, and instantly see results with helpful visualizations like word clouds, sentiment charts, and confidence scores.

The goal of this project is not only to detect misinformation but also to make the process transparent, reliable, and user-friendly. By giving people tools to verify information for themselves, we hope to raise awareness and reduce the harmful effects of misinformation in everyday life.

1.2 PROBLEM STATEMENT

In today's digital era, the rapid spread of misinformation poses a serious challenge to society. Fake news articles, misleading claims, and manipulated media such as deepfake images circulate widely across social platforms, influencing public opinion, creating panic, and sometimes leading to harmful consequences. With the growing sophistication of text generation and image manipulation tools powered by artificial intelligence, it has become increasingly difficult for individuals to distinguish between authentic and fabricated information. While several research efforts have focused on **fake news detection** using Natural Language Processing (NLP) and **deepfake detection** using Deep Learning (DL), most of these approaches address these problems separately. This lack of integration leaves a significant gap in building robust, real-world systems capable of detecting multiple forms of misinformation together.

The problem becomes more severe considering that misinformation does not exist in a single form. A fake article may contain manipulated claims, while an accompanying image or video may be synthetically generated, further reinforcing the false narrative. The absence of a **multi-modal detection system** makes it easier for misinformation to bypass detection filters, spread rapidly, and cause widespread impact. Moreover, the need for **real-time claim verification** through fact-checking services is still underdeveloped in current solutions. Therefore, there is an urgent requirement for a **unified platform** that can simultaneously analyze text for fake news, evaluate images for deepfakes, and integrate fact-checking APIs for validating claims against trusted sources. Such a system would not only strengthen digital media literacy but also provide an effective tool to mitigate the societal risks of misinformation.

1.3 OBJECTIVES

1.3.1 To design and develop a machine learning framework for misinformation detection by combining Natural Language Processing (NLP) for fake news detection and Deep Learning models (Vision Transformer) for deepfake image analysis.

1.3.2 To provide real-time claim verification by integrating the Google FastCheck API and validating news content against trusted fact-checking sources.

1.3.3 To collect, preprocess, and integrates multi-model datasets (textual news articles, claims, and images) to create a unified system capable of handling both text-based and image-based misinformation.

1.3.4 To make all components that are available in interactive web applications using Streamlit and Flask.

1.4 SIGNIFICANCE AND MOTIVATION OF THE PROJECT WORK

Misinformation has become one of the biggest challenges of our digital era. With the rapid rise of social media platforms and online news outlets, false information—whether in the form of fake news articles or manipulated images—spreads faster than ever before. This not only misleads the public but also threatens trust in media, fuels social unrest, and even impacts democratic decision-making.

The motivation behind this project comes from the urgent need to create a reliable system that can detect misinformation in both text and image formats. Traditional methods often focus on just one type of data—either news text or images—but misinformation today is multi-dimensional. By combining **Natural Language Processing (NLP)** for detecting fake news with **Vision Transformers** for analyzing deepfake images, this project aims to build a more comprehensive and effective framework.

The significance of this work lies in its potential to contribute to a safer digital environment. By enabling **real-time claim verification** through trusted fact-checking sources and integrating diverse datasets, this system can serve as a practical tool to help individuals, media organizations, and policymakers combat the spread of false information. Ultimately, the project is motivated by the vision of creating a digital ecosystem where information can be trusted, and truth has a stronger voice than misinformation.

Moreover, this project is not only academically significant but also practically relevant in today's world. As misinformation continues to evolve in sophistication, our system can serve as a foundation for future research and real-world applications. It has the potential to be integrated into news portals, social media platforms, and fact-checking tools, making it a scalable solution with long-term impact. By combining advanced AI with societal responsibility, the project aspires to bridge the gap between technology and trust in the digital age.

1.4.1 Significance of the project:

1. Provides a multi-modal detection system that can analyze both text (fake news) and images (deepfakes), making it more reliable than single-source approaches.

2. Enhances trust in media and online content by offering real-time claim verification against authentic fact-checking sources.

3. Can be used by individuals, journalists, and policymakers to quickly identify and counter false information.

4. Contributes to building a safer and more informed digital environment by reducing the harmful effects of misinformation.

1.4.2 Motivation for the project

1. The increasing spread of fake news and deepfakes inspired the need for a strong solution that addresses both simultaneously.

2. Traditional detection systems usually focus only on text or images, which motivated us to create a unified framework that handles both.

3. The misuse of advanced AI tools for generating realistic fake images and misleading content highlights the urgency to develop AI-powered countermeasures.

4. The project is driven by the vision of protecting truth and reliability in today’s digital communication space.
5. A strong motivation comes from the social impact—helping people make better-informed decisions by ensuring they can trust the information they consume.

1.5 ORGANIZATION OF PROJECT REPORT

This report is organized into 3 main chapters, with each section providing a detailed description of different aspects of the project work.

1.5.1 CHAPTER 01:

1. **Introduction:** Provides background information on the growing issue of misinformation in the digital era, highlighting the spread of fake news and deepfake images, and introduces the concept of an AI-driven system for detection.

2. **Problem Statement:** Defines the key problems the project aims to address, including the limitations of existing detection methods and the urgent need for a unified, multi-modal approach.

3. **Objectives:** Outlines the main goals of the project, such as detecting fake news using NLP, analyzing deepfake images with Vision Transformers, integrating Google FactCheck API for real-time claim verification, and combining multi-modal datasets.

4. **Significance and Motivation of the Project Work:** Explains the practical importance of tackling misinformation, its societal impact, and the motivation behind building a system that enhances trust in digital information.

5. **Organization of Project Report:** Provides a roadmap of the chapters and sections.

1.5.2 CHAPTER 02:

1. **Overview of Relevant Literature:** Reviews existing work on fake news detection, deepfake image analysis, and fact-checking systems, offering insights into different machine learning and deep learning models applied in this field.

2. **Key Gaps in the Literature:** Identifies shortcomings in current solutions, such as their limited ability to process both text and images simultaneously, lack of real-time claim verification, and dataset constraints. These gaps provide opportunities for innovation in this project.

1.5.3 CHAPTER 03:

1. **Requirements and Analysis:** Specifies the requirements of the system including tools, libraries, APIs, and datasets, along with functional and non-functional specifications.

2. **Project Design and Architecture:** Explains the system architecture, describing how NLP, Vision Transformers, and Google FactCheck API are integrated into a unified misinformation detection framework.

3. **Implementation:** Provides details on the implementation process, including dataset collection and pre-processing, training models, and integrating different modules into the complete system.

1.6 Technical Requirements (Hardware)

1.6.1 Development Environment

1. **CPU:** A modern quad-core or octa-core processor with a clock speed of at least **3.0GHz** is required to handle compiling, debugging, and running multiple tools smoothly.

2. **Ram:** At least **16GB** of RAM is recommended so that development tools, IDEs, and testing applications can run together without performance **issues.**

3. **Storage:** Either **HDD or SSD** can be used, but **SSD** is preferred for speed. A minimum of 500GB free space is needed for projects, dependencies, and related files.

4. **OS:** Compatible with **macOS, Windows, or Linux**, depending on developer preference and project requirements.

1.6.2 Production Environment

1. **CPU:** Quad-core or dual-core processor with a clock speed of at least 2.0GHz.

2. **Ram:** At least **4GB RAM** is required to ensure stable operation and smooth handling of applications.

3. **Storage: HDD or SSD** with at least **100GB free space** to store application files, logs, and updates.

4. **OS:** Supports **macOS, Windows, or Linux** as per deployment requirements.

5. **Network:** Minimum bandwidth of **10–15 Mbps** is required for reliable connectivity and stable production performance.

Chapter 02: Feasibility Study, Requirements Analysis and Design

2.1 Feasibility Study

Deepfake-News Shield's implementation is achievable given the availability of inexpensive and open-source machine-learning frameworks, natural language processing (NLP) services, and image forensics datasets, such as FaceForensics++. Technically, previously implemented models like convolutional neural networks (CNN), recurrent neural networks (RNN), and transformers can be successfully applied to detect fake articles or media that are manipulated. Economically, the use of open-source libraries as well as academic resources the cost to implement Deepfake-News Shield will be significantly reduced. Socially, Deepfake-News Shield is very relevant and covers issues of public concern, namely, the rising prevalence of misinformation and the manipulation of media that is damaging or harmful to society, political parties, politicians, etc.

Table 2.1: Literature Survey

S.No.	Title	Work done	Pros	Cons
1.	*A Comprehensive Survey on Fake News Detection Using Machine Learning (2021)* [1]	Reviews ML and DL methods for fake news detection, covering NLP-based feature extraction, contextual cues, datasets, and evaluation metrics.	Broad overview, multimodal focus, identifies gaps.	Only survey, no implementation, lacks validation.

2.	*FaceForensics++: Learning to Detect Manipulated Facial Images (2019)* [2]	Introduces FaceForensics++ dataset with 1.8M manipulated images, proposes a benchmark, and develops a CNN-based detection pipeline.	Large dataset, standard benchmark, high detection.	Face-only, drops under compression, costly training.
3.	*Multimodal Fake News Detection: A Survey of Text and Visual Content Integration Methods (2022)* [3]	Surveys multimodal detection combining text and images using deep learning, explains feature extraction and fusion strategies.	Covers multimodal, fusion boosts accuracy.	Costly model, weak generalization
4.	*A robust ensemble model for Deepfake detection of GAN-generated images on social media (2021)* [4]	Proposed VOTSTACK, an ensemble using Decision Tree, Logistic Regression, and SVM with Voting + Stacking, plus PCA-based preprocessing, achieving ~91.6% accuracy.	High accuracy, robust, scalable.	Heavy computation, not real-time.
5.	*Robust manipulated media localization and detection based on high frequency and texture features (2022)* [5]	Introduced RMLD-HFTF, combining frequency + texture features with attention and encoder-decoder structure for detection and localization of manipulations.	Detects & localizes edits, cross-dataset robust.	Complex model, high resource needs.

6.	*Deepfake Image Detection using Vision Transformer Models (2023)* [6]	Implemented a Vision Transformer (ViT) on 40,000 Kaggle images (20k real, 20k fake) to classify deepfakes, achieving 89.91% accuracy. Compared ViT performance with other detection methods.	High accuracy, fast convergence, scalable for large inputs, robust detection.	Needs large datasets, overfits on small data, computationally heavy, limited generalization.
7.	*Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms (2021)* [7]	Developed a fake news detection system using ML (LR, SVM, DT, NB), DL (LSTM), and BERT on 26k news articles. BERT achieved the highest accuracy of 98%.	Uses multiple models, strong preprocessing, BERT reached 98% accuracy.	Dataset-limited, resource-demanding, weaker ML models underperform.
8.	*Enhancing Deepfake Detection: A Multimodal Approach for Improved Accuracy (2022)* [8]	Proposed a multimodal deepfake detector leveraging blur, residual noise, and facial warping artifacts. Combined visual, noise, and landmark cues for better accuracy.	Multimodal features improve accuracy, temporal inconsistencies captured, robust detection.	Complex system, requires large datasets, high computational cost.
9.	*The New Paradigm of Deepfake Detection at the Text Level (2020)* [9]	Explored ML/DL methods for deepfake detection using CNNs, RNNs, and preprocessing techniques. Evaluated models on public datasets with accuracy, precision, recall, and F1-score.	Covers multiple ML/DL approaches, strong preprocessing, good benchmark comparisons.	Dependent on dataset quality, high computation, limited generalization to new manipulations.

10.	*Deepfake detection using deep learning methods: A systematic and comprehensive review (2021)* [10]	Conducted a survey of DL-based deepfake detection across images, videos, and audio. Reviewed datasets, key models (CNN, RNN, GAN, hybrids), and highlighted challenges and research gaps.	Comprehensive coverage, clear taxonomy, advanced techniques explained, identifies gaps.	Focused only on recent works, DL-heavy bias, requires large datasets, limited real-world validation.
11.	*Deepfake Detection Challenge Dataset (2020)* [11]	Created the DFDC dataset using diverse generation techniques (GAN, Deepfake, non-learned) and organized a benchmark challenge for detection models.	Large diverse dataset, boosted research on detection algorithms.	Weak generalization, models rely on dataset artifacts.
12.	*Deep Fake: An Overview (2021)* [12]	Reviewed deepfake techniques and security risks; proposed ECC + DNA-based encryption for securing IoT devices.	Highlights security threats, efficient ECC-based solution.	Focus on security not detection, weak generalization, vulnerable to attacks.
13.	*DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection (2020)* [13]	Surveyed face manipulation methods (identity, expression, attribute) and datasets like FaceForensics++ & DeepFakeTIMIT, summarizing key detection approaches.	Clear taxonomy, useful reference for researchers.	Uses older datasets, limited multimodal coverage.

14.	*GazeForensics: DeepFake Detection via Gaze-guided Spatial Inconsistency Learning (2023)* [14]	Introduced eye-gaze inconsistency as a deepfake cue, combining gaze estimation with CNN features; tested on FaceForensics++ and DFDC datasets.	Novel biometric-based method, better than baseline CNNs.	Drops on compressed videos, dataset-specific limitations.
15.	*Detection of Fake News Using Deep Neural Networks (2022)* [15]	Compared DNN, LSTM, BERT, and Hybrid LSTM-BERT for fake news detection on SBFN dataset, finding BERT most effective.	Strong text-based detection, BERT shows high accuracy.	Text-only focus, ignores non-textual features, limited scope.
16.	*The Emergence of Deepfake Technology: A Review (2019)* [16]	Analyzed 84 articles (2018–2019) covering deepfake uses, threats, and countermeasures across politics, business, and society.	Highlights positive applications in entertainment, healthcare, education, advertising, and AI innovation.	Exposes risks to democracy, media trust, and security, enabling disinformation and cybercrime.
17.	*SpotFake: A Multi-modal Framework for Fake News Detection (2019)* [17]	Proposed SpotFake, combining BERT for text and VGG-19 for images, tested on Twitter and Weibo datasets.	Standalone classifier, simpler design, outperforms prior models on multiple datasets.	Limited on long articles, fusion method is basic, leaves scope for improvement.

18.	*Fake News Detection After LLM Laundering: Measurement and Explanation (2025)* [18]	Evaluated fake news detectors against LLM-paraphrased text, analyzing weaknesses and detection failures.	Comprehensive study, explains failures via sentiment shifts, released useful datasets.	Detectors weaker on LLM fakes, Pegasus hardest to detect, LLM-based detectors struggle with self-generated text.
19.	*Enhanced deepfake detection with DenseNet and Cross-ViT (2025)* [19]	Introduced hybrid DenseNet+Cross-ViT with a voting mechanism for multi-face detection in videos.	Achieved near-perfect AUC and F1 on DeepForensics 1.0 and strong results on CelebDF.	Training time is high, resource-intensive, generalization to unseen deepfakes uncertain.
20.	*DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection (2020)* [20]	Surveyed face manipulation techniques and detection, categorizing synthesis, identity swap, attribute, and expression manipulation.	Useful for industries like gaming, films, cosmetics, and 3D modeling.	Enables misinformation, fake news, identity fraud, and harmful content like fake profiles or porn.

2.2 Problem Definition

In the current digital age, the proliferation of misinformation and manipulated media presents a significant risk to individuals, organizations, and society as a whole. Fake news articles can wrongly inform readers, sway public opinion, and disrupt social cohesion. Similarly, manipulated photographs and videos (deepfaking) are utilized to defame individuals, con people, or share propaganda.

Traditional approaches to fact-checking and verification of media are largely manual, and therefore slow, and not equipped to cope with the sheer volume of content available online. Additionally, any form of detection often fails when faced with cleverly executed deep-learning manipulations or deceptive narratives in text.

This emphasizes the critical need for an intelligent, automated, and scalable solution to be able to detect false news articles as well as manipulated images quickly and in real-time in order to protect digital trust and secure the responsible sharing of fact-based information.

2.3 Problem Analysis

1. **The rapid spread of false information:** Social media platforms spread false information at extremely high speeds, reaching millions of users before fact-checkers can act.

2. **Advances in deepfake technology:** Sophisticated AI techniques can produce extremely realistic-looking images/videos that are difficult for humans to spot.

3. **Limitations of manual fact-checking:** Human fact-checking is slow and laborious and cannot keep pace with the amount of information on the internet.

4. **Challenges of datasets:** There are few reliable, large-scale datasets of both fake/real news and fake images.

5. **Cross-domain limitations:** Detection processes trained to detect misinformation in a particular language, culture, or on a platform often do not translate to others.

6. **Trust and interpretability:** AI models should be used as "black boxes" that on the premise of little or no explainability can make predictions.

2.4 Solutions

To tackle these issues, we introduce Deepfake-News Shield, an innovative machine learning mechanism that can discover fake news articles and altered images. This proposed system will combine Natural Language Processing (NLP) and Computer Vision (CV) technologies to present a comprehensive problem-solving measure for multimedia misinformation.

Some key features include:

1. **Fake News Detection (Textual Analysis):** Implementation of NLP techniques (TF-IDF, embeddings, transformers) which will be applied to examining the article's content, writing style, and the credibility of its source.

2. **Manipulated Image Detection (Visual Analysis**): Utilization of CNN-based deep learning models trained on datasets like FaceForensics++ for photo manipulation detection.

3. **Aggregate ML Models**: Usage of ensemble learning combined with deep learning to improve the overall accuracy in terms of detecting text and images as fake.

4. **Explainability:** Added interpretable AI methods to focus in on suspicious words, phrases, or image regions to build user trust.

5. **Centralized Dashboard:** Simple interface to provide real-time monitoring with detection reports.

6. **Scalability and Automation:** The proposed system is designed for the collection of a copious amount of social media/article collections, and to adapt to new manipulative behaviors.

2.5 Requirements

2.5.1 Functional Requirements

1. **Data Collection & Preprocessing:** To establish a diverse dataset, the system will acquire news articles and images from reliable and unreliable sources. All text data will be cleaned, tokenized, and normalized, while all image data will be resized and augmented to enhance robustness and accuracy.

2. **Fake News Detection Module:** The system will employ supervised ML models, such as Logistic Regression, Random Forest, and BERT, on

labeled datasets (real news and fake news). The module will ideally classify articles as real articles or fake articles with high accuracy.

3. **Manipulated Image Detection Module:** Image data will be analyzed using CNN architectures, specifically XceptionNet and ResNet. The system will classify images as original images or manipulated images, where applicable, and indicate whether or not the original images have been tampered with.

4. **Ensemble and Hybrid Integration:** The predictions made from the text and image modules will be combined using ensemble methods such as voting and stacking to increase reliability and robustness.

5. **User Dashboard:** A simple dashboard will give real-time results along with confidence scores alongside summaries of articles. Users will also be able to upload articles or images for verification by the system. The user dashboard will maintain a log of previous article and image detection activities made by the system for future review.

6. **Reporting and Analytics:** The outcomes of all detections made by the system will be retained for future analytic endeavors. The system will provide reporting on detection performance indicating the systems overall accuracy levels, types of errors, and trends of detected articles or images for future improvements.

2.5.2 Non-Functional Requirements

1. **Scalability:** The ability to accommodate sizable datasets and real-time inputs from various sources.

2. **Accuracy & Reliability:** Achieve high precision and recall in fake detection, even with noise/compression.

3. **Real-Time Responsiveness:** Provide low-latency detection for quick-moving social media activities.

4. **Explainability:** Supply interpretable results to increase user confidence in predictions made by the system.

5. **Maintainability:** Modular architecture that allows the model to be easily retrained through new datasets.

6. **Security & Privacy:** Secure sensitive user-uploaded data and ensure the ethical use of the detection results.

7. **Robustness:** Tolerant to adversarial manipulation of photos, dataset bias, and evolving deepfake alterations.

2.6 Flow Chart

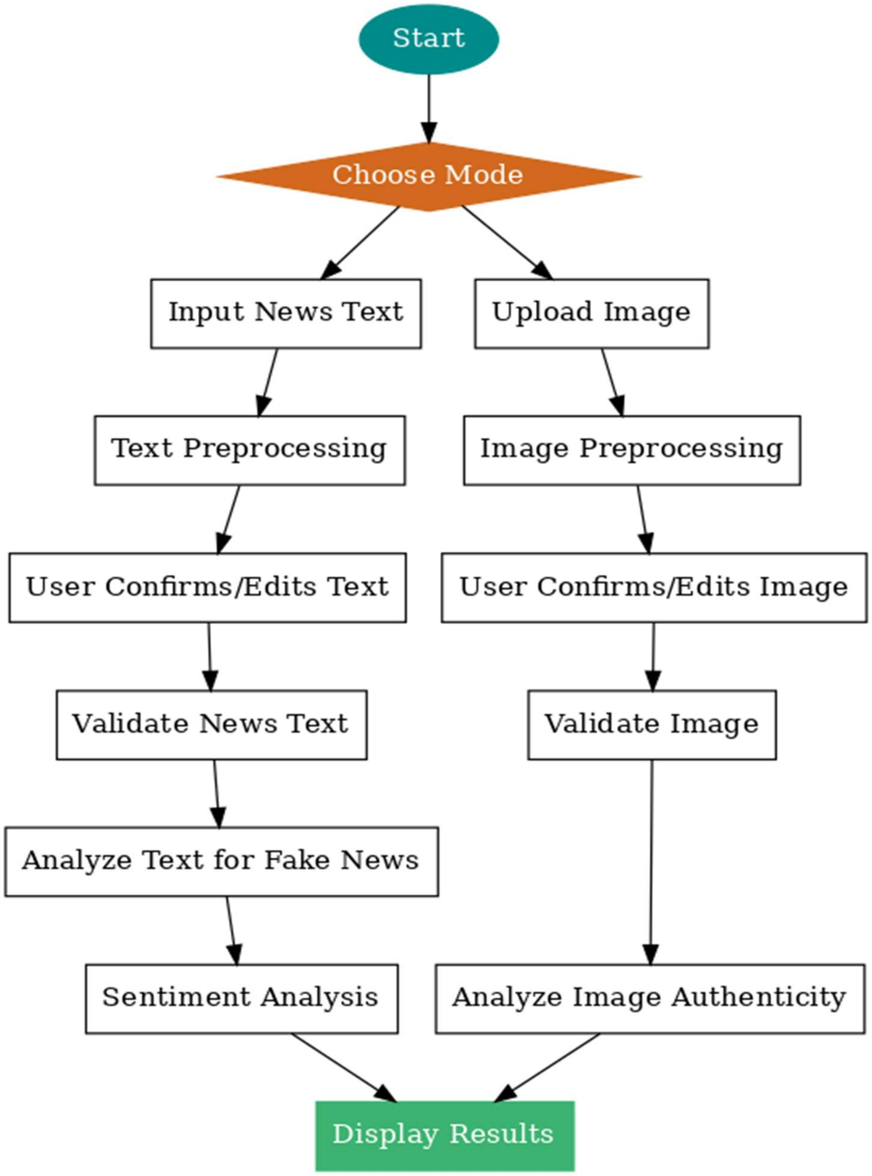
The flowchart demonstrates the functioning of the Deepfake-News Shield system. The first step in the process is for the user to select a type of input—a news story or an image.

1. In the case of the Fake News Detection path, first, the input text is preprocessed, validated, and then categorically analyzed using machine-learning models to classify the article as real or fake. Additionally, sentiment analysis can also lend more reliability in categorizations.

2. In the Manipulated Image Detection path, the uploaded image again goes through a process of preprocessing and validation, followed by analysis with deep-learning models to determine it to be authentic or manipulated.

3. Finally, both branches converge at the Display Results stage, where the system provides the outcome along with confidence scores for the user.

This ensures a unified and reliable solution for detecting both false news content and manipulated images.



[Fig1. Flow Chart of Deepfake-News Shield]

Chapter 03: System Development

3.1 Project Design and Architecture

The project Deepfake-News Shield: Detecting False Articles and Manipulated Photos is designed to address the growing challenge of misinformation and manipulated media circulating on online platforms. The system architecture combines Natural Language Processing (NLP) and Computer Vision (CV) methods to analyze both textual content and images for authenticity.

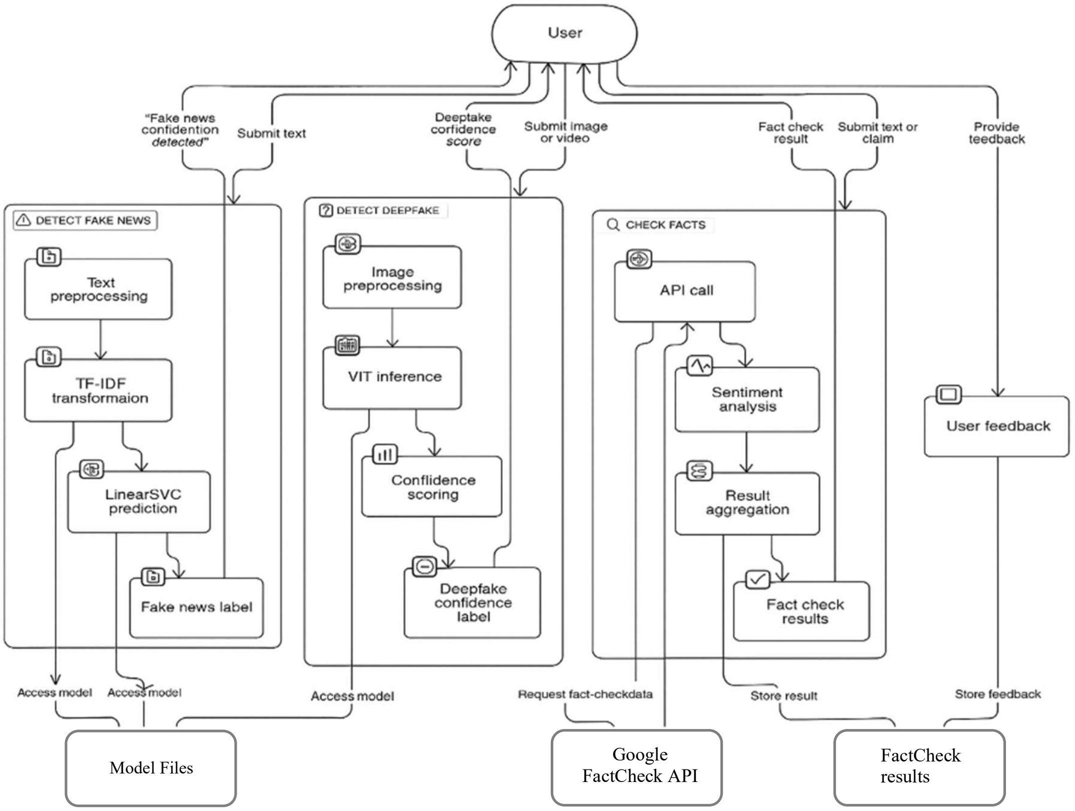
The architecture is divided into three primary modules:

1. **Fake News Detection:** Input text undergoes preprocessing (cleaning, tokenization, and feature extraction using TF-IDF), followed by classification through supervised machine learning models. The output is a prediction label indicating whether the news content is real or fake.

2. **Manipulated Face Detection:** Uploaded images are preprocessed and analyzed using deep learning-based models such as Vision Transformers (ViT) or CNN variants. The system generates a confidence score and labels the image as original or manipulated.

3. **Fact-Checking & Feedback:** A fast-checking component integrates external APIs (e.g., Google FactCheck API) to cross-verify claims, while user feedback is stored for further improvement of the system.

The overall architecture ensures **clarity, modularity, and scalability**, allowing future integration of more advanced detection models and larger datasets. The flowchart below illustrates the interaction between modules, starting from user input to final result display, emphasizing both automation and user transparency.



[Fig2. Project Design and Architecture]

3.2 Frontend Implementation

At this stage in the project, only the front-end of the system has been completed. A simple, user-friendly website was created to serve as an interface for the Deepfake-News Shield.

The homepage has two main features:

1. Input a news article in text form for verification.
2. Upload an image to check for manipulation.

The design ensures clarity and easy navigation for users. Once the input is provided, it is intended to pass through preprocessing and validation steps, as illustrated in the **flowchart** in **Chapter 2.**

3.3 Implementation Stack

The frontend of the system was created with the following technologies:

4. **HTML:** Utilized for constructing the structure and layout of the web pages.

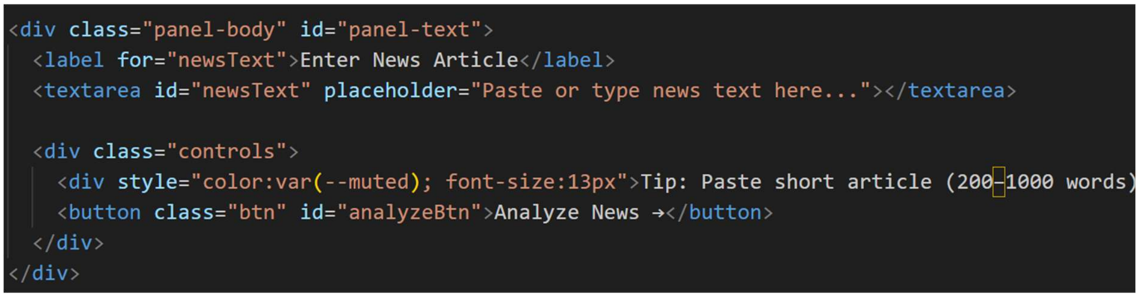
5. **CSS:** Used for the styling and visual design as well as making the interface responsive for different devices.

6. **JavaScript**: Added interactivity, enabling dynamic behaviour such as form submission and handling user inputs.

These technologies were chosen for their simplicity, compatibility, and efficiency in creating lightweight and responsive web applications.

3.4 Code Snippets

1. HTML *(Snippet)*

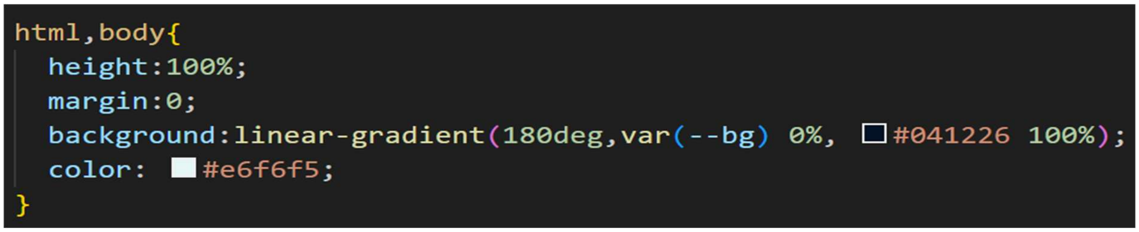


```html
<div class="panel-body" id="panel-text">
  <label for="newsText">Enter News Article</label>
  <textarea id="newsText" placeholder="Paste or type news text here..."></textarea>

  <div class="controls">
    <div style="color:var(--muted); font-size:13px">Tip: Paste short article (200–1000 words)
    <button class="btn" id="analyzeBtn">Analyze News →</button>
  </div>
</div>
```

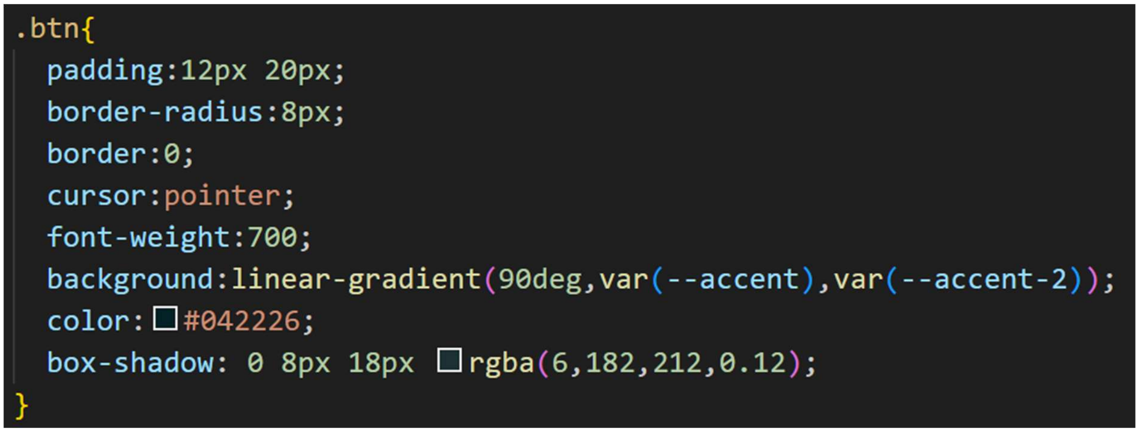
[Fig3. This snippet creates the main input area where users can paste a news article and click the button to analyze it.]

2. CSS *(Snippet)*



```css
html,body{
   height:100%;
   margin:0;
   background:linear-gradient(180deg,var(--bg) 0%, ■#041226 100%);
   color: □#e6f6f5;
}
```

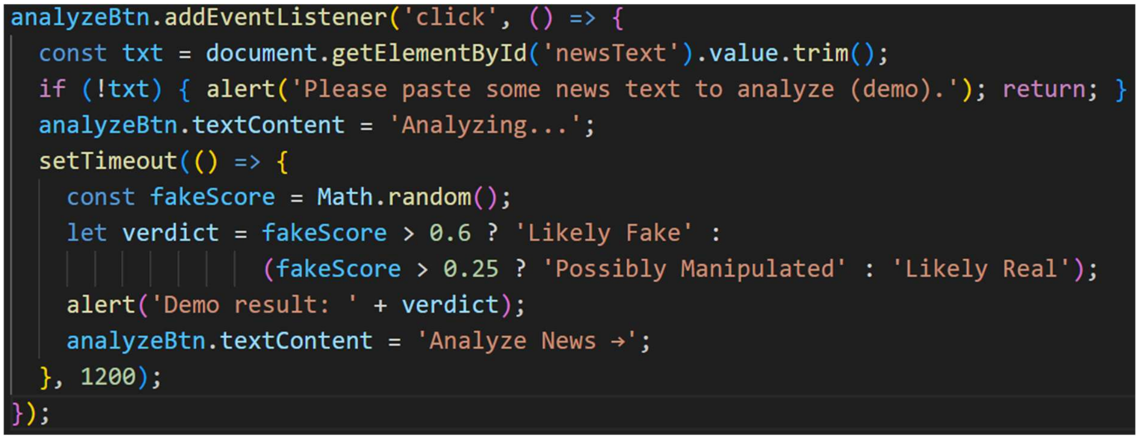
[Fig4. This snippet defines the dark gradient background and global text styling for the website.]



```
.btn{
  padding:12px 20px;
  border-radius:8px;
  border:0;
  cursor:pointer;
  font-weight:700;
  background:linear-gradient(90deg,var(--accent),var(--accent-2));
  color:■#042226;
  box-shadow: 0 8px 18px ■rgba(6,182,212,0.12);
}
```

[Fig5. This snippet creates the modern gradient buttons used for actions like “Analyze News.”]

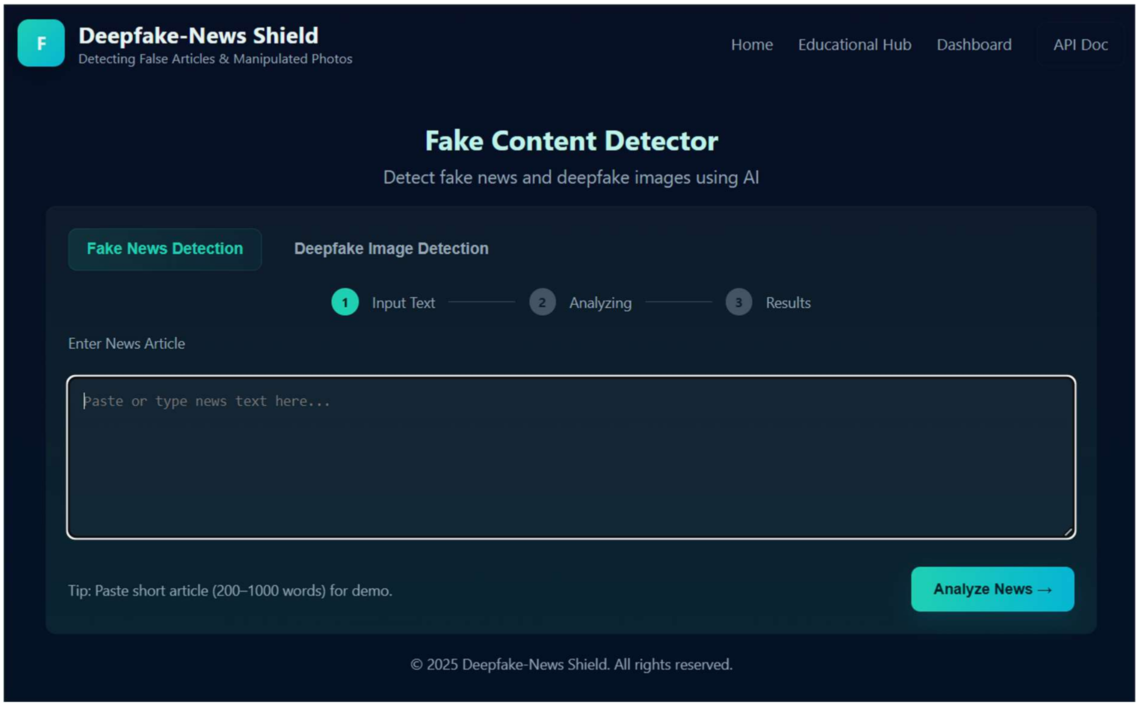
3. JS *(Snippet)*



```
analyzeBtn.addEventListener('click', () => {
  const txt = document.getElementById('newsText').value.trim();
  if (!txt) { alert('Please paste some news text to analyze (demo).'); return; }
  analyzeBtn.textContent = 'Analyzing...';
  setTimeout(() => {
    const fakeScore = Math.random();
    let verdict = fakeScore > 0.6 ? 'Likely Fake' :
                  (fakeScore > 0.25 ? 'Possibly Manipulated' : 'Likely Real');
    alert('Demo result: ' + verdict);
    analyzeBtn.textContent = 'Analyze News →';
  }, 1200);
});
```

[Fig6. This snippet shows the core logic for analyzing news text in **demo mode**, randomly generating a fake/real result.]

3.5 Result



Detect fake news and deepfake images using AI

Tip: Paste short article (200–1000 words) for demo.

© 2025 Deepfake-News Shield. All rights reserved.

[Fig7. Website's Home Page]

3.6 Future Work

1. **Integration of a Backend:** As of now, the system contains elements of front-end functionality. Ongoing work will involve implementing a back-end and integrating it to provide real-time analysis of both text inputs and photo inputs.

2. **Deployment of Machine Learning Models:** Deep learning models (for example, BERT will be used for the text, and XceptionNet or ResNet may be employed for image processing) will be developed and subsequently deployed. Subsequently, detection methods will classify news articles and/or images with improved accuracy as either fake--manipulated--or real, as part of overall framework.

3. **Expansion of Datasets:** To enhance model robustness and generalization, larger and more diverse datasets of fake news articles and manipulated image data will be gathered.

4. **Highlighting Manipulated Regions:** In the case of image manipulation detection, the system will be enhanced to classify images as being fake or real, and highlight or identify manipulated or altered regions, if feasible, in order to enhance transparency.

5. **Real-Time Dashboard & Analytics:** An additional user dashboard will be developed that will allow the user to view detection results, store the detection results, and analyze statistics about detection accuracy, output, or trends.

REFERENCES

[1] S. Kaliyar, A. Goswami, and P. Narang, “FakeBERT: Fake news detection in social media with a BERT-based deep learning approach,” *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11765–11788, 2021.

[2] M. Kaur, S. Kumar, and S. Bawa, “Rumor detection on social media: A data mining perspective,” *Information Systems Frontiers*, vol. 24, no. 5, pp. 1485–1506, 2022.

[3] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “FaceForensics++: Learning to detect manipulated facial images,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1–11.

[4] H. Nguyen, J. Yamagishi, and I. Echizen, “Use of a capsule network to detect fake images and videos,” *arXiv preprint* arXiv:1910.12467, 2019.

[5] M. H. A. Khan and F. Algarni, “Deepfake detection using convolutional neural networks for improved security on social media,” *IEEE Access*, vol. 9, pp. 143825–143837, 2021.

[6] B. Ghita, I. Kuzminykh, A. Usama, T. Bakhshi, and J. Marchang, “Deepfake Image Detection using Vision Transformer Models,” IEEE/academic publication, 2023–2024

[7] N. N. Prachi, M. Habibullah, M. E. H. Rafi, E. Alam, and R. Khan, “Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms,” Journal of Advances in Information Technology, vol. 13, no. 6, pp. 652–661, Dec. 2022, doi: 10.12720/jait.13.6.652-661.

[8] Karthikeyan A., Monniesh B., Kishorekumar V., and Niveshkumar S., “Enhancing Deepfake Detection: A Multimodal Approach for Improved Accuracy,” TIJER – International Research Journal, vol. 11, no. 7, July 2024, pp. 583–587.

[9] C.-M. Rosca, A. Stancu, and E. M. Iovanovici, “The New Paradigm of Deepfake Detection at the Text Level,” Appl. Sci., vol. 15, art. 2560, Feb. 27, 2025, doi: 10.3390/app15052560.

[10] A. Heidari, N. J. Navimipour, H. Dag, and M. Unal, “Deepfake detection using deep learning methods: A systematic and comprehensive review,” WIREs Data Mining and Knowledge Discovery, vol. 14, no. 2, Feb. 2024, Art. no. e1520.

[11] N. Shakya and P. Poudyal, "Detection of Fake News Using Deep Neural Networks," Kathmandu University Journal of Science, Engineering and Technology, vol. 16, no. 2, pp. 110–119, 2022.

[12] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," Information Fusion (Elsevier), 2020.

[13] Z. He, et al., "GazeForensics: DeepFake Detection via Gaze-guided Spatial Inconsistency Learning," arXiv preprint, 2023.

[14] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The DeepFake Detection Challenge (DFDC) Dataset," arXiv preprint arXiv:2006.07397, 2020.

[15] B. B. Gupta, S. Kumar, Shubham, and A. Jaiswal, "Deep Fake: An Overview," in Smart and Innovative Trends in Engineering and Technology (SITET-2020), 2021.

[16] M. Westerlund, "The Emergence of Deepfake Technology: A Review," Technology Innovation Management Review, vol. 9, no. 11, pp. 39–52, Nov. 2019.

[17] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, "SpotFake: A Multi-modal Framework for Fake News Detection," in Proc. IEEE Int. Conf. Multimedia Big Data (BigMM), 2019.

[18] R. K. Das, "Fake News Detection After LLM Laundering: Measurement and Explanation," arXiv preprint arXiv:2501.18649, 2025.

[19] F. Siddiqui, J. Yang, S. Xiao, and M. Fahad, "Enhanced deepfake detection with DenseNet and Cross-ViT," Expert Systems With Applications, vol. 267, p. 126150, 2025.

[120] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," arXiv preprint arXiv:2001.00179, 2020.