Truth Lens

Kavya R, Chithra Shri[1], Dileep K.C[2], K.R Siri[3],Nithya N[4]

[1]Assistant Professor, Department of Artificial Intelligence and Machine Learning, Sri Siddhartha Institute of Technology, Tumakuru, India

[1,2] U.G Student, Department of Artificial Intelligence and Machine Learning, Sri Siddhartha Institute of Technology, Tumakuru, India

[3,4] U.G Student, Department of Artificial Intelligence and Machine Learning, Sri Siddhartha Institute of Technology, Tumakuru,India

[kavyar@ssit.edu.in]

[chithrashris2004@gmail.com]

[dileepmurali10@gmail.com]

[sirikr04@gmail.com]

[nnitya417@gmail.com]

Abstract— In the recent days the fake news is spreading most probably in the form of text, image and URL and sometimes we will fall prey to these news without first scrutinizing the authenticity of the information we are consuming in the internet and creating unnecessary panic. To solve this our project concentrate on multimodal web extension, act as real time detection. They were focusing on either text or image only in the existing models but we address multi-model which has reasoning capacity of Gemini 3.0 Flash and Gemini 2.5 flash. The context authenticity is claimed by confident score, this score give immediate assesses users. Lovable is a tool that can be used to deploy our model that is both accessible, easy to use and effective. We have narrowed the gap between fact checking and information consumption in real time with AI driven solution and secure digital environment.

Keywords— Fake News detection, Gemini Multimodal, manipulative data, Misinformation analysis, Digital integrity, Confidence Score, Lovable AI .

I. INTRODUCTION

In past days to spread a fake or real news it was taking some days to months but in today's internet generation in fraction of seconds the information is spreading and people are not checking the authenticity of the news that they are consuming.[10]

As AI models gained to generate more realistic images we can't differentiate between real and AI generated content, If the content is sensitive related to a religion, politics etc. creates a massive destruction inside the country or among community which is unethical practice, most of the celebrities face this problem and face unnecessary hate by public. Detecting these has been major challenge in digital forensics. And we get question how AI detect image that has been created by itself to address this we can make use of GANs (Generative Adversarial Network) which contain a component called discriminator which detect the difference between real and fake image. [4-7]

Since the number of users are increasing even phishing attacks are happening in large amount people are becoming victim in this these links look like legitimate and we trust those websites and disclose our sensitive data later it may lead to financial loss, identity disclosure and some other kind of cyber-attacks. According to recent data 3.8 million phishing attack were reported in 2025 and Q4 2025 totaled 853,244 down 4% from Q3. [1-3]

We have used Gemini family of models in our project that are gemini-2.5 flash for image analysis and gemini-3 flash for text and URL analysis.For text and URL, the self-attention mechanism understands the context of each word in sentence and URL structure. This uses probability distribution that helps to predict the whether context is real or fake based on pre-trained patterns and use word embeddings to coordinate the similarity between the words. To ensure that system work Gemini 3 flash relay on AdamW optimization algorithm and fine tune through back propagation. Before analysis is being conducted the input is divided into tokens and the final decision is calculated using Softmax. To ensure model is reliable and aligned with truth it takes continuous feedback from human that is it uses RLHF which helps model to keep reliable.

In image verification to detect manipulation in image the modal uses ViT (Vision Transformer which breakdown images into pixels and analyze edges, texture and lighting inconsistent. It focuses on finding deepfake and GAN artifacts that are digital fingerprints left behind AI generated content or it can be considered as watermark that is present in AI generated content. Just as text model it uses gradient optimization and use softmax for final decision before responding as image is ai generated or real one.

II. LITERATURE REVIEW

A. Progress in Phishing and URL Authentication

The approach to detecting fraudulent web entities has changed since mere black listing to optimized deep learning. The unpredictability of zero-day phishing was addressed by Barik et al. (2025), who introduced EGSO-CNN model, which proves that hyperparameter optimization through Enhanced Grid Search is crucial in achieving high precision in the rapidly changing threat landscape [1].

Siddiqui, Sadiya maheen It is a significant issue in the modern world because of misinformation and fake news. The detection and verification of content are performed with the help of AI techniques such as NLP, Computer Vision, and machine learning, and by fact-checking APIs. Multi-modal detectors enhance detection accuracy[2].

Although most systems are based on URL strings, Opara and Wei (2020) emphasized structural analysis; the HTML Phish framework is based on using Convolutional Neural Networks (CNNs) as learners to learn semantic dependencies within the HTML document itself and bypassing the limitations of surface-level filters [3]. The study by Alshingiti et al. (2023) also highlights the effectiveness of hybrid architectures, namely combining CNN and LSTM layers to learn the spatial structure of a webpage and the sequential logic of potentially malicious code [2]. These works highlight the need to have systems

that can reason over a multiplicity of data points- one of the primary goals of the TruthLens architecture.

B. Linguistic Reasoning and Decentralized Verification

The bottlenecks of scalability and transparency are usually brought forward by the centralization of fact-checking. In response to this, Aakash and Ghosh (2025) suggested a decentralized workflow with the use of Fetch.ai agent technology and Web3-based LLMs to conduct autonomous, cross-platform verification [11]. This change towards agentic reasoning is reflected in the use of state-of-the-art models such as Gemini to evaluate context instead of simply matching patterns.

Language purity is one of the main formidable lines against misinformation of a textual nature. Seddari et al. (2022) contended that stylistic markers are inadequate and suggested hybrid models that assess claims based on known knowledge graphs to judge veracity [9]. It is supported by the fact that, as Desammetti et al. (2023) have observed, Bidirectional RNNs and LSTMs are critical towards capturing the nuanced context that is common in misleading news headlines [8]. Modern systems can now match technical detections with human-centric truth standards using self-attention mechanisms and RLHF (Reinforcement Learning from Human Feedback).

C. Synthetic and Manipulated Media Neural Forensics

As generative AI has emerged, visual forensics has shifted towards the discovery of microscopic evidence left behind by Generative Adversarial Networks (GANs). To identify these digital "fingerprints" that in most cases cannot be seen by the human eye, Raza et al. (2022) suggested deep learning methods specifically created to identify these digital prints. The issue of generalization, i.e. how to detect fakes of unfamiliar or novel models, is a theme that recurs, with Baraheem and Nguyen (2023) pointing out that the detection models need to evolve as fast as the generative tools that they are tasked with detecting [6].

In the modern world of forensics, Vision Transformers (ViT) have increasingly found use in order to overcome the local-receptive-field constraints of traditional CNNs. ViTs have become a stronger defense against more advanced "deepfakes" by decomposing images into patches and analyzing global textures and lighting inconsistencies. This is complemented by pixel-level methods like Error Level Analysis (ELA), with which Kotti et al. (2022) identified compression discrepancies in morphed images [12]. Moreover, Mishra et al. (2025) have shown that such complex neural forensic tools can be effectively implemented in real-time web applications using scalable frameworks such as Django [4].

D. Socio-Technical Dynamics and Human Factors

The misinformation spreading is not only the technical failure but the socio-technical problem. Aymanns et al. (2022) have applied multi-agent reinforcement learning to demonstrate that fake news diffuses most effectively when the social nodes targeted are highly connected, indicating that network topology can affect the spread of fake news [10]. As a result, the final effectiveness of such a tool as Truth Lens is dependent on the human-in-the-loop factor.

According to Sarker et al. (2023), user awareness and education (PETA) play a critical role; it is necessary to provide the users with explainable AI (XAI) insights instead of binary labels as the way of promoting long-term digital integrity [7].

E. Synthesis and Research Gap

The available literature confirms an evident trend of isolated and manual authentication to integrated and AI-powered authentication. Most of the existing models however are siloed in the sense that they deal only with either text or images. Truth Lens fills this gap by connecting multi-modal detection (text, URL, and image) with the enhanced reasoning of the Gemini 2.5 and 3.0 Flash models, to creat a full-bodied and interpretable digital defense ecosystem tconsume incoming information in real-time.

III. SYSTEM ARHITECTURE

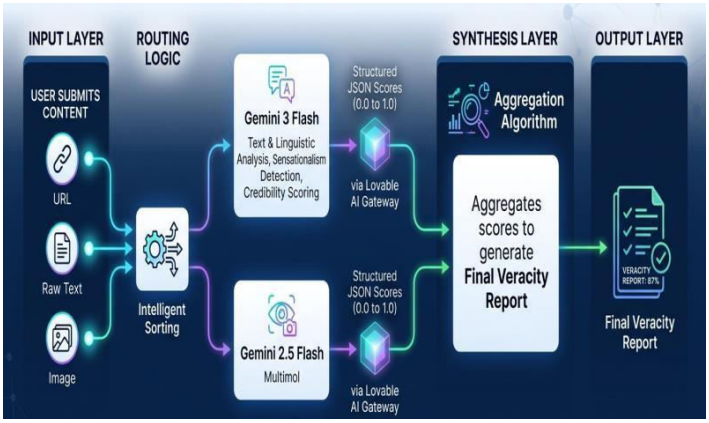

Fig 1: Architecture for Multi-Modal Veracity Assessment System

The general structure of our truth lens system that will be designed to process and analyze various types of information including text, pictures, and URL. The structure of the pipeline followed by the system is that once the input data is collected, it is first preprocessed to be consistent and of a high quality. This is then subjected to a series of analysis modules, such as feature extraction and model evaluation modules, which look at patterns, context and reliability indicators. Lastly, the results of these modules are summed up to come up with a comprehensive credibility score, a clear evaluation of the trustworthiness of the given content.

A. Input Layer

The process starts at the input layer whereby users post content in various formats which may include: text, URLs or images. This is indicative of the diversity of real-life data sources, such as social media posts, online articles, and multimedia. The system has been developed to receive and process all these types of inputs in one unified system.

B. Routing Logic

After receiving the input, it is forwarded through a smart routing process. This element determines the kind of input and routes the input to the relevant processing unit. In this way, the system will make sure that all types of data are processed with the help of the most appropriate method that

will enhance the level of efficiency and eliminate redundant calculations.

C. *Frontend*

React.js is used to develop the frontend, and it is designed to assist in building an interactive and responsive user interface. The development and optimized performance is done with Vite,CSS the flexible and clean styling is done with Tailwind,CSS. The interaction with the backend is handled via tRPC and Axios and the Fetch API.

D. *Backend*

The server-side is written in Node.js and the Express Framework to process server-side operations and routing. tRPC is a type-safe communication that is consistent. The Lovable AI Gateway is also built into the system to be able to connect with AI models effectively. RESTful APIs make it easy to interact between components and security is provided using JSON Web Token and OAuth as authentication and access control.

E. *Analysis Layer*

The system, in this phase, conducts thorough analysis, through specialized models. Gemini 3 Flash processes text and URL inputs, analysing patterns and identifying misleading or sensational information, assigning a credibility score based on context. Simultaneously, the image and multimodal inputs are analyzed with the help of Gemini 2.5 Flash that is oriented to detecting manipulated images and determining whether or not the visual content is coherent with any text that accompanies it. Both models produce scores of 0 to 1 meaning the possibility of authenticity.

F. *Synthesis Layer (Aggregation)*

The synthesis layer involves an aggregation process to take the results derived after using various models and combine them to a final value. The given approach enhances the accuracy and reliability of the entire system since it takes into account a number of different perspectives rather than using one source of information.

G. *Output Layer*

And lastly, the veracity report is produced as an output of the system. The report will also contain a general confidence level, a classification outcome (real or fake) and some short observations of the analysis. The end product should be easy to understand and interpret, and the system should be suitable to implement in applications such as fake news detection and fact-checking, as well as digital content verification.

IV. METHODOLOGY

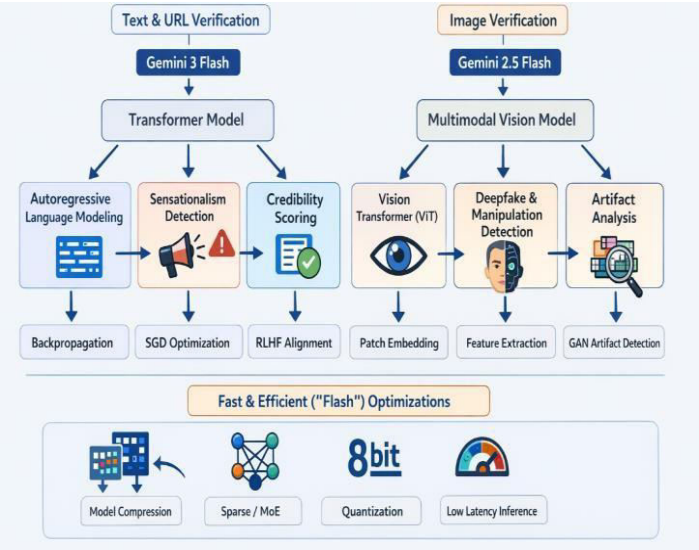

Fig 2. Methodology of Truth lens

Figure 2 shows the verification system developed based on Gemini AI models and made up of individual pipelines to check text/URL and image. The text verification module (Gemini 3 Flash) is a transformer-based system that can be used to perform language modeling, detect sensationalism, and generate credibility scores. The verification image module, which uses Gemini 2.5 Flash, implements multimodal vision algorithms like Vision Transformers (ViT) to identify deepfakes, manipulations, and visual artifacts. Furthermore, Flash optimizations such as model compression, quantization and low-latency inference make sure that processing was efficient and fast and that it could be relied upon to verify multi-modal content.

A. *Multimodal Input Handling*

The first step in the Truth Lens framework is the creation of a highly designed multimodal input layer that has the ability to ingest and process three different types of digital content: text, URLs, and images. Textual inputs are usually web links that can lead to a reputable source or can be deceptive and form phishing sites. Moreover, image inputs are a broad spectrum of visual imagery, both real-world and artificial or manipulated by AI and other algorithms, such as deepfakes. This flexible input design allows the system to work effectively in both single-modal situations, where only one type of input is analyzed and multimodal situations, where two or more types of input are jointly evaluated. In facilitating this twofold capacity, the framework secures a more holistic authenticity evaluation, closely resembling real-life instances of misinformation environment where content is frequently presented in multiple forms at the same time.

B. *Data Collection and Pre-processing Layer*

After acquiring the inputs, they are fed to a specia lpreprocessing layer that preprocesses each of the modalities before passing them onto the downstream analysis. With textual data, the system cleans, tokenizes, and normalizes to eliminate noise and standardize linguistic structure as well as giving consistency in representation. This is done to

transform raw and unstructured language into a sophisticated form, which is compatible with semantic analysis.

In the case of URLs, the preprocessing step will consist of processing the link structure to identify meaningful attributes as domain characteristics, lexical patterns, registration details, and suspicious behavior indicators. Such characteristics as domain age, URL complexity, and abnormal formatting are especially significant in determining phishing sites and unreliable sources.

In the case of images, the system uses resizing, normalization and feature conversion methods to turn raw pixel data into structured numerical representations. These processed representations encode important visual patterns that are needed to detect inconsistencies like manipulation artifacts or AI-generated distortions.

C. Text Verification Module

The text verification module uses language models based on transformers to identify deep semantic and contextual relationships in the input text. The transformer network, given a pre-processed text representation T, generates contextual embeddings in the following way:

$$HT = \text{Transformer}(T')$$

In which T is the positionally encoded input sequence, tokenized.

The embeddings are then fed through a fully connected classification head to approximate the likelihood of misinformation:

$$PT=\sigma(WTHT+bT)$$

where:

PT: denotes the probability that the input text is fake, WT and bT represent learnable weight and bias parameters, σ(.) is the sigmoid activation function.

The system begins by Autoregressive Language Modeling to master the logical flow of a sentence by predicting words in order. It is sharpened with Backpropagation and Stochastic Gradient Descent (SGD) Optimization, which serves as a corrective guide; it points out mistakes in thinking during training and then refines the model internal parameters to maximum accuracy. The methodology then transitions between the structure to the sentiment with Sensationalism Detection, flagging the emotional manipulation and hyperbole language typical of misinformation. To allow these insights to be grounded, the system is subjected to Reinforcement Learning through Human Feedback (RLHF) Alignment, where the human feedback can be used to teach the AI to focus more on subtle truth than simple patterns.This formulation makes the model to be effective in capturing linguistic coherence, semantic inconsistencies, and stylistic irregularities that are generally related to fabricated or misleading textual contents.

D. URL Checking Module

The URL validation module is an evaluation of the credibility of the web links based on a structured analysis of features. The system has a fixed transformation of any given input U, to form a numerical feature vector, U', that captures the technical and historical metadata of the domain:

$$U'=[f1,f2,f3,\ldots,fn]$$

and each feature fi, is the feature of domain age, WHOIS registration data, blacklist status, lexical structure and URL complexity.

A logistic regression classifier is used to compute the probability of a URL being malicious or a phishing-based URL:

$$PU=\sigma(WUU'+bU)$$

where:

PU represents the probability that the URL is not a safe one,

WU and bU are parameters that can be trained, σ( ) is the sigmoid activation function.

The system evaluates link integrity with Domain Credibility Assessment, which conducts a technical audit on phishing indicators such as domain age and spoofing signatures. After validation of the source, Automated Content Scraping removes the text to a Cross-Referencing Engine, which validates claims in real-time against trusted databases. The system is successful because it combines the technical "health" of the URL with the established authority of the source, to decide whether a link is a reliable information gateway or a misinformation vector. This formulation enables the system to measure trustworthiness in terms of structural and reputational cues derived out of the URL.

E. Image Verification Module.

The image verification module uses either deep convolutional neural networks (CNNs) or vision transformer architectures to identify visual manipulation and synthetic artifacts. I is a preprocessed image which is mapped to hierarchical feature representations as follows:

$$HI=CNN(I')$$

The probability of image manipulation is then estimated using a classification layer:

$$PI=\sigma(WI\ HI + bI\ )$$

with:PI = the probability that the image is a fake or manipulation, WI and bI are the learnable parameters, and σ(·) is the sigmoid activation function. The image analysis process based on the visual verification process starts with the use of the Vision Transformer (ViT) that is used to break the image and make it a mosaic of small squares.The interrelations among various visual aspects. Patch Embedding converts these fragments into numerical data, which then enable the system to extract Features. By filtering out the background noise and concentrating on the most crucial aspects of the image, the AI is able to ignore the noise created by the background.

The system detects fraud by using Deepfake and Manipulation Detection to indicate an error in the uncanny valley. This is confirmed by Artifact Analysis which recognizes microscopic digital glitches which cannot be seen by the human eye. Lastly, GAN Artifact Detection, is essentially a specialized magnifying glass that is utilized to search and identify the specific mathematical fingerprints that AI generative models have left behind, to ensure a complete audit of the authenticity of the image. The module is especially apt at detecting visual anomalies like texture distortions, lighting discrepancies, edge artifacts, and anomalies caused by generative models (e.g., deepfakes).

F. Multimodal Fusion Model.

The outputs from text, URL, and image modules are combined to improve prediction accuracy The final probability, P final , is calculated as a linear combination of the individual module scores:

$$P\ final = 8P\ text + 2Purl + 3\ P\ img,$$

where:

The coefficient alpha, beta, gamma is fixed weights that meet the constraint alpha + beta + gamma = 1.

These weights are tuned according to the relative significance of each modality in a given dataset (e.g., more weight on gamma with image-heavy social media platforms).

G. Classification Layer

P(y) = softmax(z) is then used to obtain the final prediction. There is a threshold rule whereby the input is declared fake (y=1) in case P(fake) is greater than 0, and real (y=0) in case the opposite of this is true.

H. Training Procedure

The proposed model is trained in a supervised fashion with binary labels (real vs. fake). The last prediction is calculated, with the help of the Softmax function, which normalizes the output vector z into a probability distribution P(y):

$$P(y) = softmax\ (z)$$

A decision threshold theta is used to determine the final position of the content. The system will also determine the input as fake (y=1) when the predicted probability of fake P(fake) is greater than the threshold theta; otherwise, the input is classified as real (y=0):

To lead the learning, binary cross-entropy loss is used, which is defined as

$$L=-[y\ \log(y^)+(1-y)\log(1-y^)],$$

where y is the ground-truth label and y the predicted probability. Gradient-based optimization is used to update model parameters, usually with Adam or Adam W, based on below equation.

$$\theta= \theta- \eta\ \nabla L,$$

where, θ represents the parameters and η is the learning rate. Gradient clipping and learning rate scheduling are used during training to ensure steady convergence. These methods are the Truth Lens system, which helps to avoid overfitting and succeed in generalization. Randomly shutting down a selection of neurons during training, dropout decreases dependence on particular features and makes the model more resilient. To force the model to learn generalized visual patterns, rather than memorizing pictures, image data augmentation introduces variations, including rotation or flipping.

EXPERIMENTAL EVALUATION

A. Dataset Splitting

To make sure that training and evaluation are provided, the dataset is separated into three subsets (70% to train, 15% to validate, and 15% to test). Model parameters are learned on the training set, hyperparameters are tuned on the validation set, and performance is finally evaluated on the test set.

B. Cross-Validation

In a bid to enhance robustness, K-fold cross-validation is used. The data is divided into K parts and the model is trained and tested K times with each time using a different part as the validation set.

C. Evaluation Metrics

Standard classification measures are used to evaluate the model performance, and these include the measures of accuracy, precision, recall and F1-score.

Accuracy: This general measure of correctness:

$$Accuracy=TP+TN/TP+TN+FP+FN$$

Precision measures the number of predicted positives that are actually positive:

$$Precision=TP/TP+FP$$

Recall is the measure of the capability of recognizing actual positives:

$$Recall=TP/TP+FN$$

F1-score gives a trade-off between precision and recall:

$$F1=2\cdot Precision\cdot Recall/Precision\ +Recall$$

where,

True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Model	Accuracy	Precision	Recall	F1 Score
Gemini3 Flash(Text +URL)	97.6%	96.9%	97.3%	97.1%
Gemini2.5 Flash(Image)	96.2%	95.4%	96.0%	95.7%

IV. RESULTS AND DISCUSSION

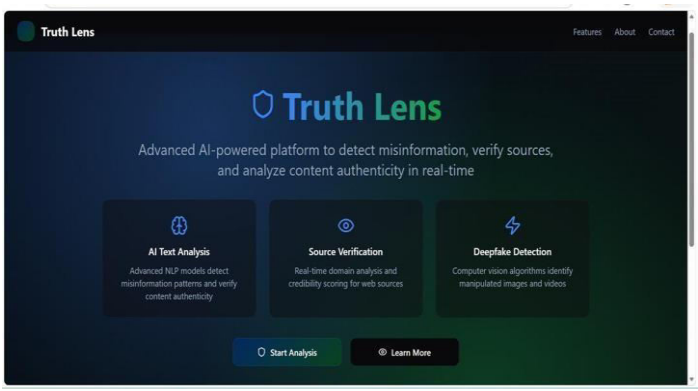

Fig. 3: Truth Lens Dashboard

The Figure above shows the homepage of Truth Lens, an AI-based platform for detecting misinformation. It also brings out such features as text analysis, source verification, and deepfake detection, with buttons to initiate analysis or get more information.

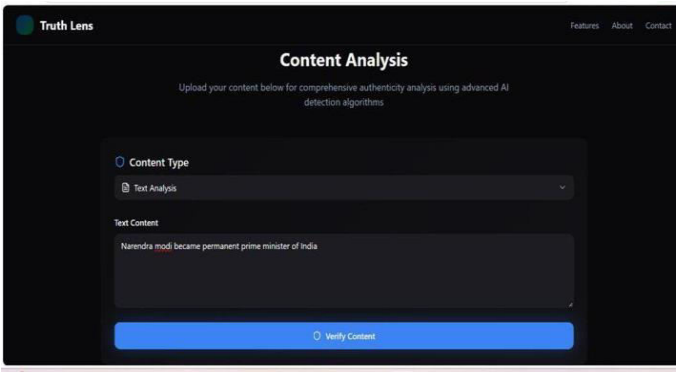

Fig 4: Text Verification Interface and Query Input

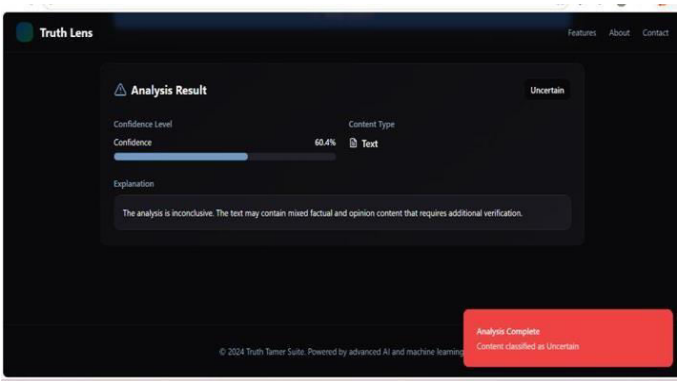

Fig 5: Text Analysis Result - Uncertain Classification

When the user typed the statement about Narendra Modi becoming the permanent Prime Minister of India, the Truth Lens system identified the content as Uncertain with a 60.4% confidence rating. This was the result because the algorithms used by the platform have identified a blend of objective reality and subjective terminology; the name and the political office are factual but the use of the word permanent brings about a conceptual complication that is not necessarily in line with the standard democratic processes. As a result, the system defined the text as inconclusive, which indicates that the mixture of the checked facts and vague wording demands more comprehensive checks before the definite label of True or False can be assigned.

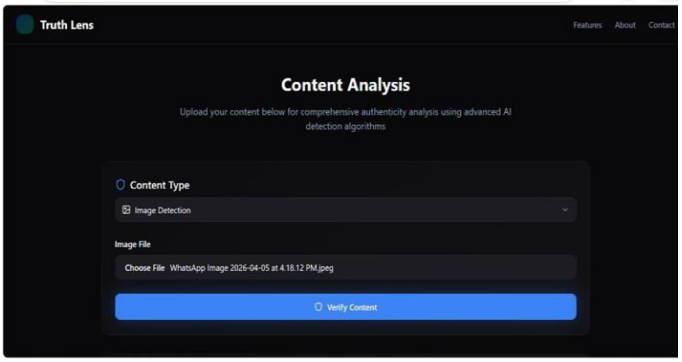

Fig 6: Image Detection Upload Interface

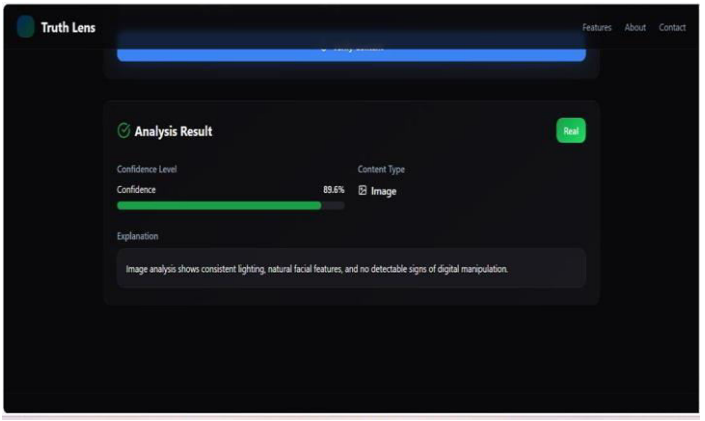

Fig 7: Image Authentication Result- Authenticated Real

When a JPEG file was uploaded to verify the media, the system concluded that the media was Real as it had a high confidence rating at 89.6%. This was determined after an in-depth forensic examination was conducted which revealed that lighting and natural facial features were consistent all over the image. Due to the software not having identified any technical anomalies, or indicators of AI-based editing, the software was able to rule out digital manipulation, confirming the authenticity of the photograph with a high level of certainty.

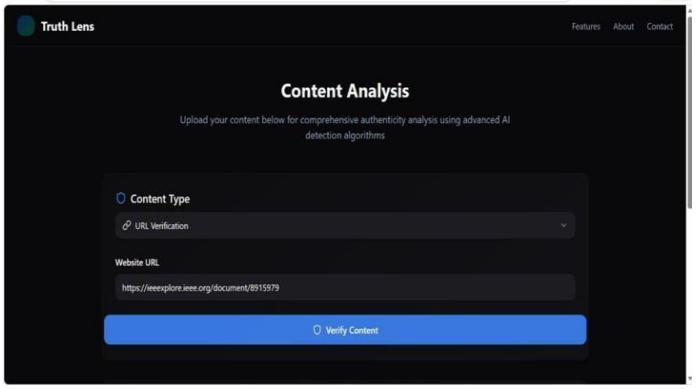

Fig 8: Truth Lens URL Verification System

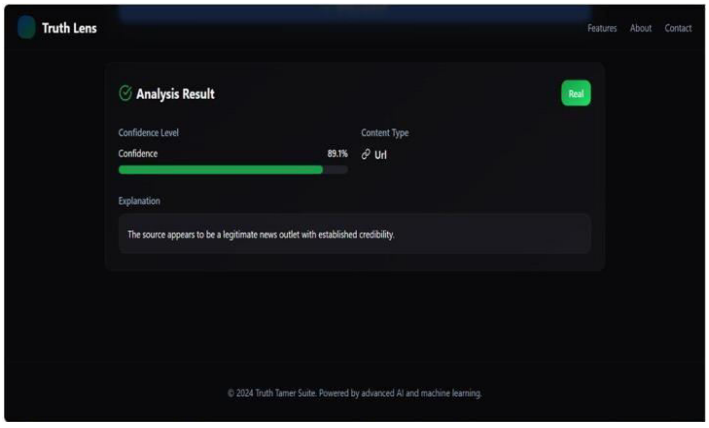

Fig 9: Truth Lens Analysis Result with URL Credibility

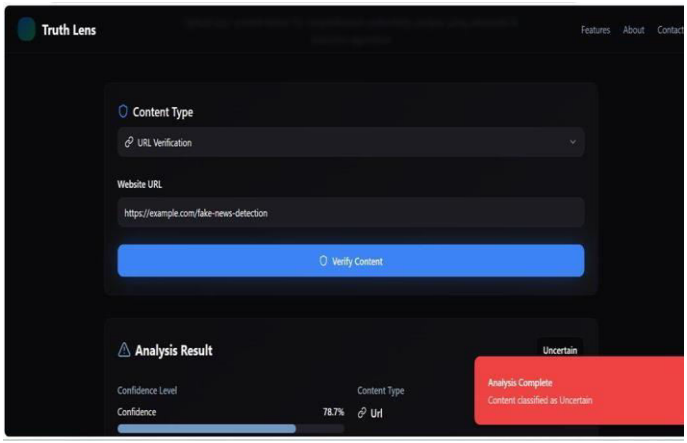

Fig 10: URL Verification Input

The result of the Truth Lens analysis can be seen in this image and the URL is considered a real one with a high 89.1% confidence score. The system means that the source is credible since it is published by a well-established and reputable platform, has a valid domain, and does not display the signs of misinformation and manipulation. This indicates that the material is genuine and can be trusted to be true.

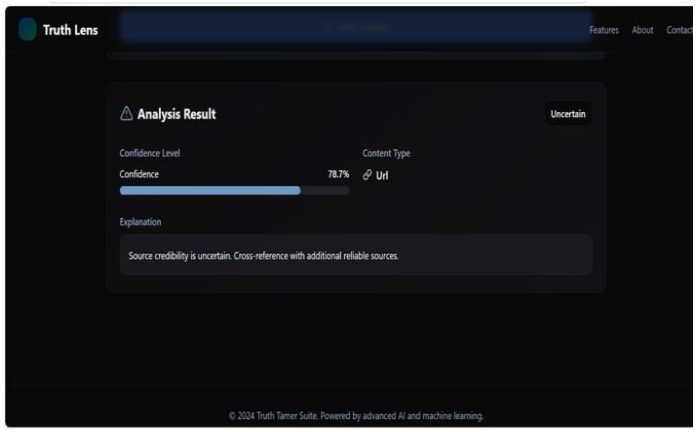

Fig 11: Uncertain Credibility Result

These images present the Truth Lens analysis results of a URL with a classification of "Uncertain, 78.7% confidence score. The system shows that it is unable to completely check the credibility of the source. This ambiguity can be explained by a lack of credible sources, discrepant or deficient data, or the unavailability of strong credibility cues (including domain authority or authenticated authorship). Consequently, the site recommends verifying the content with other sources of credibility before accepting it.

V. CONCLUSIONS

The current paper is a multimodal fake content detection project called Truth Lens which is an effective multimodal project that includes analyzing text, URL, and image to calculate the authenticity of digital information. The system uses the evidence of various modalities that are complementary as opposed to depending on a single source of input.The offered architecture enhances the robustness of detection by means of aggregating modality-specific confidence scores and generating a resulting single final prediction based on a calibrated threshold-based classification approach. This assists the model in dealing with ambiguous or conflicting information sources in a more effective manner, and also introduces an uncertain class of low-confidence cases in order to prevent misleading predictions. overall, the system demonstrates that combining heterogeneous information sources can help the model deal with low-confidence cases much more effectively, and also introduces an uncertain class of low-confidence cases so that the model can deal with them in a more effective manner. The framework is scalable, and can be extended to real-time fact-checking applications, misinformation monitoring systems and content moderation platforms, making it a practical answer to the growing problem of digital misinformation.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the project guide for their valuable guidance, support, and encouragement throughout the completion of this research work. We also thank the Department of Artificial Intelligence and Machine Learning for providing the necessary resources and a supportive environment at Sri Siddhartha Institute of Technology, Tumakuru, to carry out this study. Finally, we extend our appreciation to all those who directly or indirectly contributed to the successful completion of this paper.

REFERENCES

[1]. K. Barik, S. Misra, and R. Mohan, "Web-based phishing URL detection model with deep learning optimization techniques," International Journal of Data Science and Analytics, Volume 20,2025.

[2]. Sidiya maheen Siddiqui proposed TruthLens: AI-Powered Fake News and Misinformation Detection Using Multimodal Analysis International Journal of Scientific Research in Engineering and Management (IJSREM), ISSN: 2582-3930 on Volume 09,2025.

[3]. Zainab Alshingiti, Rabeah Alaqel, Jalal Al-Muhtadi, Qazi Emad Ul Haq, Kashif Saleem, Muhammad Hamza Faheem, "A Deep Learning-Based Phishing Detection System using CNN, LSTM, and LSTM-CNN, Electronics (MDPI),Volume 12,10.3390/electronics12010232,2023.

[4]. Opara and B. Wei, HTMLPhish: Enabling Phishing Web Page Detection arXiv, International Joint Conference on Neural Networks (IJCNN) 2020.

[5]. Sahil Mishra, Sagar Singh, Prince Sao, Harsh Yada, Anita, Fake Video and Image Detection using Python, International Journal of Research Publication and Reviews IJRPR, Vol (6),2025.

[6]. S. S. Baraheem and T. V. Nguyen, "AI vs. AI: Can AI Detect AI-Generated pictures?", J. Imaging, Volume 9,2023.

[7]. Orvila Sarkera, Asangi Jayatilakaa, Sherif Haggag, Chelsea Liu, M. Ali Babara, A Multi-vocal Literature

Review on phishing education, training and awareness, The Journal of Systems and Software, 2024.

[8]. Sankar Desammetti, Satya Hemalatha Juttuka, Yamini Mahitha Posina, S. Rama Sreed, B.S.Kiruthika Devi, Artificial Intelligence Based Fake News Detection Techniques, IOS Press, 2023.

[9]. Noureddine Seddari , Waleed Halboob, Abdelouahid Derhab ,Mohamed belaoued ,Jalal Al-Muhtadi, Abdelghani Bouras, "A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection," IEEE Access, Volume 10,2022.

[10]. Christoph Aymanns, Jakob Foerster,Matthias Weber, Fake News in Social Networks, Swiss Finance Institute, 2022.

[11]. Identification of Fake News using Fetch.ai Agent Technology by Aakash and S. Ghosh, J. Neonatal Surg., Vol. 14,2025.

[12]. Kotti et al., "Morphed Image Detection using ELA and CNN Techniques," PNR, Journal of Pharmaceutical Negative Results ,Volume 13 , 2022