

DeepTruth AI: A Multi-Modal Deepfake and Fake News Detection

Ms. Shreya Sunil Shinde¹, H. R. Vyawahare²

¹Student, Computer Science and Engineering, Sipna College of Engineering and Technology, Amravati

²Guide / Co-author, Computer Science and Engineering,
Sipna College of Engineering and Technology, Amravati

ABSTRACT

The Deepfakes and fake news are increasingly threatening digital trust by spreading manipulated media and misinformation. This paper presents DeepTruth AI, a multi-modal detection framework that integrates textual, visual, and audio data analysis using advanced AI models such as BERT, CNN, and LSTM. The system employs a Transformer-based fusion module and an explainability engine (ATEX) to provide accurate and transparent detection results. The framework addresses evolving challenges in misinformation and offers applications in media verification, cybersecurity, and legal forensics.

I. INTRODUCTION

Social media and digital platforms accelerate the spread of information but also facilitate misinformation, including fake news and deepfakes. Deepfakes use AI to produce synthetic images, videos, and audio that impersonate individuals or fabricate events, damaging credibility.

Fake news manipulates textual content and can influence public opinion negatively. Detecting such content is challenging due to the sophistication of generation methods. DeepTruth AI proposes a multi-modal approach combining text, image, and audio analysis to improve detection accuracy and reliability.

To address these challenges, the need for robust and intelligent detection systems has become critical. Traditional detection methods, which rely on single-modal analysis, often fail against complex and realistic manipulations. DeepTruth AI proposes a comprehensive multi-modal approach that integrates text, image, and audio analysis to identify manipulated or synthetic content with greater accuracy and reliability. By leveraging transformer models, neural networks, and cross-modal fusion techniques, the framework enhances detection performance and provides deeper insights into the authenticity of digital media.

II. LITERATURE SURVEY

Deepfake detection research widely employs convolutional neural networks (CNNs), attention-based models, and temporal analysis for video verification [1]-[3]. Blockchain-based methods embedding cryptographic fingerprints have been explored for media provenance [4]. Ethical and philosophical considerations around synthetic media's dual-use potential have also been discussed [5]. Multi-modal fusion methods combining text, image, and audio features have shown promise in enhancing detection robustness [7], yet face challenges in real-time deployment and explainability. Additionally, the rapid evolution of generative models such as GANs and diffusion models continues to outpace traditional detection techniques, making adaptive and scalable frameworks essential.

III. METHODOLOGY

The DeepTruth AI framework integrates four main methodologies:

Text-Visual DeepFusion Network (TVD-Fusion): Combines BERT embeddings from text with CNN-extracted visual features, fused using transformer-based attention binary classification (real or fake).

Audio-Visual Synchrony Detection (AVSync-Detect): Uses MFCC and LSTM for audio, facial landmark extraction for visual input, and cross-modal synchrony checks to detect lip-sync mismatches.

Lightweight Transformer (LiteT- Detector): A compact multimodal transformer employing DistilBERT and MobileNetV2 optimized for mobile/web deployment.

ANIMATED-TRUTH EXPLAINABLE

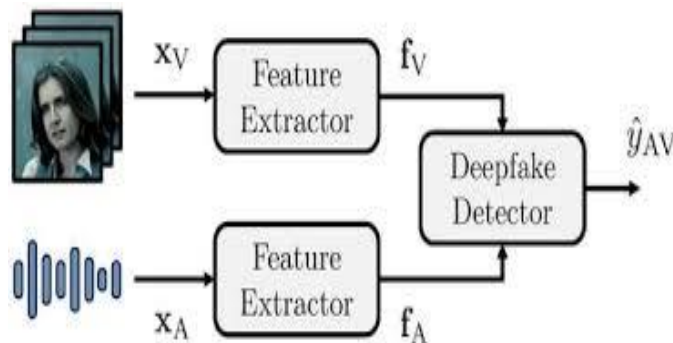
Engine (ATEX): Provides visual and textual explanations through attention maps, LIME heatmaps, and audio feature attributions to enhance transparency.

IV. ARCHITECTURE

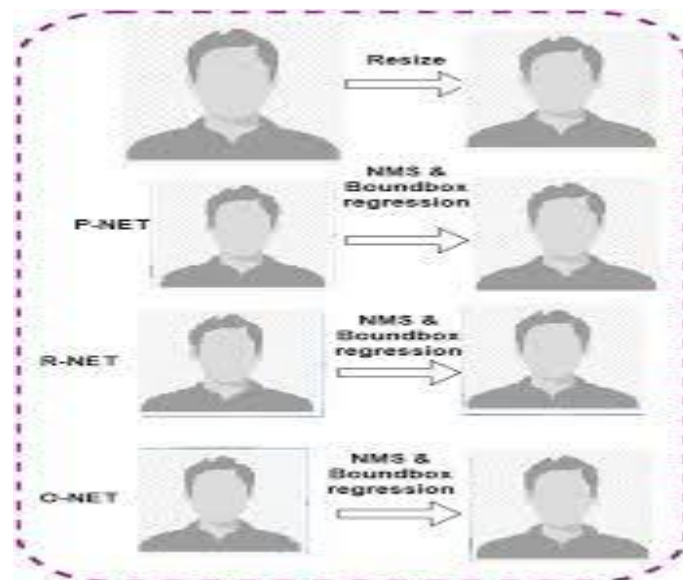
The system architecture processes multi- modal inputs:

- Textual data encoded by BERT.
- Visual data encoded by CNN (ResNet or EfficientNet).
- Audio data processed via LSTM or CNN-RNN hybrids.

The fusion layer utilizes a transformer- based attention



mechanism to integrate modalities, followed by a classification head applying softmax or sigmoid functions to output real, manipulated, or synthetic labels. Optional temporal streams analyze micro-expressions and biometric cues for enhanced liveness detection.



V. ANALYSIS OF PROBLEM

Face spoofing attacks include 2D/3D static and dynamic techniques such as printed photos, video replays, masks, and robotic faces. Deepfake-based spoofing uses AI-generated synthetic videos causing serious security risks in biometric systems.

Challenges include sensor-level spoofing, environmental exploits, and adversarial generative attacks, necessitating multi-modal liveness detection techniques.

VI. APPLICATION

DeepTruth AI is applicable in:

- Fake News Detection: Assisting media outlets in verifying content authenticity.
- Legal Evidence Verification: Validating audio visual evidence integrity.
- Cybersecurity: Enhancing biometric authentication and fraud prevention.
- Educational Tools: Raising awareness about misinformation and synthetic media.

VII. CHALLENGES

Key challenges include:

- Synchronizing multi-modal data streams accurately.
- Generalizing across diverse and unseen fake content.
- Balancing real-time performance with detection accuracy.
- Enhancing explainability for non-technical users.
- Addressing data scarcity and evolving deepfake generation method.
- Ensuring privacy and ethical compliance in biometric data handling.

VIII. FUTURE SCOPE

Future work includes:

- Social media platform integration for real-time deepfake detection.
- Multilingual and cross-cultural model enhancements.
- Adversarial training against evolving deepfake algorithms.
- Improved explainability with interactive, user-friendly tools.

- Privacy-preserving AI utilizing federated learning.
- Legal tools for certified forensic reporting.

IX. CONCLUSION

DeepTruth AI offers a scalable, accurate, and explainable multi-modal framework to tackle the rising threat of deepfakes and fake news. By fusing text, image, and audio analysis with explainability, it supports trustworthy digital content verification. Continued refinement and deployment will strengthen global to uphold information integrity in the digital age.

X. REFERENCES

1. Y. Li, M.-C. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking,” IEEE Int. Workshop Inf. Forensics Secur., 2018.
2. A. Afchar et al., “MesoNet: a Compact Facial Video Forgery Detection Network,” IEEE Int. Workshop Inf. Forensics Secur., 2018.
3. D. Guera and E. J. Delp, “Deepfake Video Detection Using Recurrent Neural Networks,” IEEE AVSS, 2018.
4. MIT DeepTruth White Paper, Cryptographic Fingerprinting for Synthetic Media.
5. Ethical Perspectives on Synthetic Media, Journal of Media Ethics.
6. R. Zhang et al., “Multi-modal Fusion Based Deepfake Detection via Semi- supervised Learning,” CVPR Workshops, 2021.