International Research Journal of Education and Technology
Peer Reviewed Journal, ISSN 2581-7795
Impact Factor 5.007

Deep Truth AI: A Multi-Modal Deep fake and Fake News Detection

Ms. S T Renuka, Mr.S. Alagu@Sivadharsan
III B.Sc. Digital and Cyber Forensic Science
Dr T.Ramaprabha, Associate Professor
Department of Digital and Cyber Forensic Science
Nehru Arts and Science College
Coimbatore

Abstract: Deepfakes and fake news are increasingly threatening digital trust by spreading manipulated media and misinformation. This paper presents Deep Truth 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. Furthermore, the proposed model leverages cross-modal consistency analysis to identify subtle discrepancies across different data streams, improving robustness against sophisticated synthetic content. Extensive evaluation demonstrates enhanced detection accuracy and generalization across diverse datasets. The inclusion of explainable outputs enables stakeholders to interpret model decisions with greater confidence. The architecture is designed to be scalable for real-time deployment across digital platforms. Overall, Deep Truth AI contributes to strengthening information integrity and fostering a secure digital communication environment.

Key words: Deepfake, Misinformation, Multimodal, Forensics, Authentication, Neural Networks, Detection

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

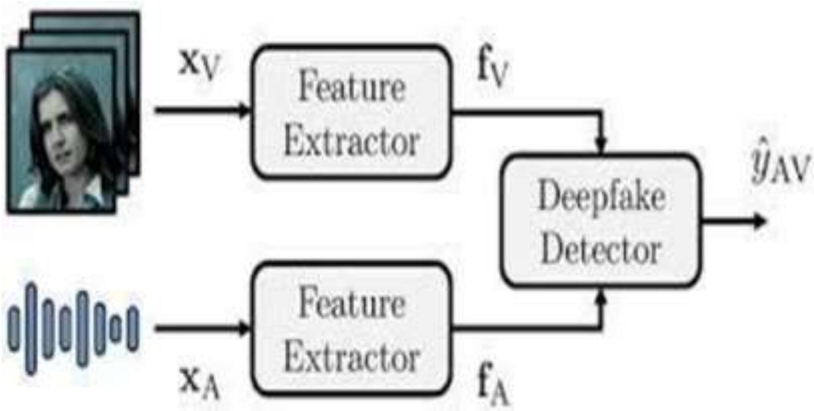
The Deep Truth AI framework integrates four main methodologies:
Text-Visual Deep Fusion Network (TVD-Fusion): Combines BERT embeddings from text with CNN- extracted visual features, fused using transformer-based attention binary classification (real or fake).
Audio-Visual Synchrony Detection (AV Sync-Detect): Uses MFCC and LSTM for audio, facial landmark extraction for visual input, and cross-modal synchrony checks to detect lip-sync mismatches.
Lightweight Transformer (Lite T- Detector): A compact multimodal transformer employing Distil BERT 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 (Res Net or Efficient Net).
- 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 soft max 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. Advanced attackers often exploit high-resolution displays and projection systems to bypass traditional face recognition algorithms. Variations in lighting conditions, camera quality, and background noise further complicate reliable spoof detection. Generative adversarial networks (GANs) continuously improve the realism of synthetic faces, making detection increasingly difficult. To counter these threats, modern systems integrate texture analysis, depth sensing, and behavioral biometrics such as eye blinking and micro-expressions. Combining hardware-based safeguards with AI-driven anomaly detection is essential to ensure robust and secure biometric authentication systems.

VI. APPLICATION

Deep Truth AI has wide-ranging applications across multiple critical domains where trust and authenticity are essential. In fake news detection, it assists media organizations and fact-checking agencies by automatically analyzing text, images, audio, and videos to verify the authenticity of digital content before publication, thereby reducing the spread of misinformation. In the legal domain, Deep Truth AI plays a vital role in evidence verification by validating the integrity of audio-visual materials presented in courts, helping ensure that manipulated or fabricated evidence is identified. From a cybersecurity perspective, it strengthens biometric authentication systems by detecting spoofed or synthetic media used in identity fraud and social engineering attacks. Additionally, Deep Truth AI enhances fraud prevention by identifying deepfake-based impersonation attempts in digital transactions. In the education sector, it serves as an effective awareness tool by demonstrating how misinformation and synthetic media are created and detected. This helps students and researchers develop critical digital literacy skills. Overall, Deep Truth AI contributes to building a more secure, transparent, and trustworthy digital ecosystem.

VII. CHALLENGES

Deep Truth AI faces several critical challenges that must be addressed to ensure robust and reliable performance. One major difficulty lies in accurately synchronizing multi-modal data streams such as text, audio, images, and video, as inconsistencies between modalities can affect detection accuracy. Another challenge is generalizing detection models to handle diverse and previously unseen deepfake techniques, since generative models continuously evolve. Maintaining a balance between real-time processing speed and high detection accuracy is also essential, especially for large-scale social media monitoring systems. Furthermore, enhancing explainability remains crucial so that non-technical users, legal professionals, and policymakers can understand and trust the system's decisions. Data scarcity poses another limitation, particularly the availability of high-quality, labeled deepfake datasets for training. The rapid advancement of deepfake generation methods further complicates model adaptation and long-term reliability. Additionally, ensuring privacy and ethical compliance while handling sensitive biometric data is a significant concern. Robust governance frameworks and secure data management practices are therefore necessary. Addressing these challenges is vital for building a trustworthy and sustainable multi-modal deepfake detection system.

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.

VIII. CONCLUSION

Deep Truth AI presents a scalable, accurate, and explainable multi-modal framework designed to combat the escalating threat of deepfakes and fake news in the digital ecosystem. By integrating advanced text analytics, image forensics, and audio signal processing, the system performs comprehensive cross-modal verification to detect inconsistencies and manipulation. Its multi-layered architecture enables robust performance across diverse platforms, including social media, news networks, and digital communication systems. The incorporation of explainable AI techniques ensures that detection outcomes are transparent and interpretable, thereby enhancing user trust and institutional adoption. This explainability feature is particularly valuable for journalists, legal authorities, and policymakers who require clear evidence to support decision-making. The framework's scalability allows it to handle large volumes of data in real-time environments without compromising accuracy. Moreover, its adaptive learning mechanisms help it respond to evolving deepfake generation technologies. By promoting reliable content authentication, Deep Truth AI strengthens cybersecurity and digital resilience. Continued research, optimization, and global deployment of such systems will significantly contribute to combating misinformation. Ultimately, Deep Truth AI plays a vital role in upholding information integrity and trust in the digital age.

VII. 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.
7. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1820.
8. Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-stream neural networks for tampered face detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
9. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22–36.
10. Dolhansky, B., et al. (2020). The DeepFake Detection Challenge (DFDC) dataset. *arXiv preprint arXiv:2006.07397*.
11. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. In *Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS)*.

 S. Alagu Sivadarsan (USN: 23UGDF032) is an aspiring professional with a strong interest in Digital and Cyber Forensic Science. He is passionate about exploring advanced techniques in cybercrime investigation, digital evidence analysis, and emerging security technologies. With a keen focus on cybersecurity, forensic methodologies, and ethical digital practices, he aims to contribute to building a secure and trustworthy digital environment. He can be reached at sivadharsan1911@gmail.com or via mobile at 9342473976 for academic and research collaborations.

 S. T. Renuka (USN: 23UGDF032) is a dedicated student with a strong interest in Digital and Cyber Forensic Science. She is keenly focused on understanding cybercrime investigation techniques, digital evidence examination, and modern cybersecurity practices. Her academic interests include network forensics, ethical hacking, and data security, with the goal of contributing to the advancement of secure digital ecosystems. She is enthusiastic about research and practical applications in the field of cyber forensics. She can be contacted at strenukamano@gmail.com or via mobile at 9043773495 for academic and professional collaborations.