iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.77220

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

A Study on Human-Computer Interaction in Deepfake Detection Systems

Miss. Renuka Khade[1], Miss. Aishwarya Dharmadhikari[2], Miss. Priyanka Khandagale[3], Prof. Rupali Saraf[4]
[1,2,3]*Department of Computer Application, K.B. Joshi Institute of Information Technology, Pune SNDT Women's University*
[4]*Asst. Professor, K.B. Joshi Institute of Information Technology, Pune SNDT Women's University*

Abstract: ***Human-Computer Interaction (HCI) plays a crucial role in the design of intelligent systems that humans can understand, trust, and effectively use. With the rapid advancement of artificial intelligence, deepfakes have emerged as a significant challenge to digital trust. Deepfakes are synthetic media generated using machine learning techniques such as Generative Adversarial Networks (GANs), capable of producing highly realistic manipulated images, videos, and audio.***
Although existing deepfake detection systems achieve high ac- curacy under controlled conditions, many prioritize algorithmic performance while neglecting usability, transparency, and human trust. This paper explores the integration of HCI principles into deepfake detection systems. We review existing detection methodologies, analyze cognitive load and explainable artificial intelligence (XAI), and examine ethical implications associated with synthetic media. A user-centered framework is proposed to enhance system usability and trust without compromising detection accuracy. The results highlight the importance of ex- plainability, minimal cognitive load, and ethical interface design in improving real-world adoption of deepfake detection tools.
Index Terms: Human-Computer Interaction, Deepfake Detection, Explainable AI, Cognitive Load, Trust in Technology

I. INTRODUCTION

Human-Computer Interaction (HCI) focuses on designing interactive systems that are usable, efficient, and aligned with human cognitive capabilities. Traditionally, HCI research emphasized interface usability, accessibility, and ergonomics. However, the growing influence of artificial intelligence has introduced complex challenges related to transparency, trust, and ethical responsibility. One such challenge is the proliferation of deep- fakes—synthetic media created using machine learning models such as GANs. While deepfakes enable innovation in entertainment and education, they also facilitate misinformation, identity manipulation, and erosion of trust in digital communication. Deepfake detection has therefore become a critical research area in computer vision and multimedia forensics. Despite advancements in detection accuracy, many systems fail to consider how users perceive and interact with detection tools. This gap limits real-world adoption. Integrating HCI principles can bridge this gap by improving usability, explainability, and ethical acceptance.

II. LITERATURE REVIEW

Early deepfake detection techniques focused on identifying visual artifacts such as facial landmark inconsistencies and abnormal blinking patterns. Frequency-domain approaches later revealed subtle spectral traces introduced during media synthesis. Recent research has adopted multimodal detection methods that combine audio, video, and temporal cues. However, studies report significant performance decay when detection models are evaluated on unseen datasets or novel deepfake generation techniques. Research on human per- formance in deepfake detection further indicates that users struggle to identify manipulated content without assistance, emphasizing the need for well-designed detection interfaces. Explainable AI (XAI) has been proposed as a solution to im- prove transparency. Prior studies demonstrate that explanations significantly enhance user trust, though excessive technical detail may increase cognitive load.

III. HCI- FRAMEWORK

The proposed framework integrates Human-Computer In- teraction (HCI) principles into deepfake detection systems to improve usability, explainability, and trustworthiness. Unlike conventional approaches that focus purely on detection accu- racy, this framework emphasizes the **user experience** and **decision-making support**. The framework consists of four primary components:

A. User-Centered Interface Design

The interface is tailored to diverse user groups, including journalists, educators, content moderators, and general social media

users. Key design considerations include:

- Adaptive dashboards: Interfaces adjust the complexity of information based on the user's expertise, offering summary-level results for novices and detailed forensic data for experts.
- Visual cues: Highlighting suspicious regions of images or video frames using heatmaps, bounding boxes, or color- coded alerts.
- Interactive controls: Users can toggle between views (e.g., frame-by-frame analysis, audio waveform inspec- tion) to better understand the detection process.

B. Explainable AI Integration

Explainable AI (XAI) enhances transparency by providing interpretable insights into model decisions. The framework incorporates:

- Feature attribution: Techniques like Grad-CAM or LIME highlight which facial regions, audio segments, or frame sequences influenced the detection result.
- Confidence scores with context: Probabilities are ac- companied by textual explanations, e.g., "This frame has a 92% likelihood of manipulation due to inconsistent eye blinking and facial symmetry anomalies."
- Decision rationale tracking: Users can view a stepwise breakdown of detection decisions, fostering trust and reducing skepticism.

C. Cognitive Load Optimization

Excessive information can overwhelm users, reducing trust and usability. The framework minimizes cognitive load through:

- Progressive disclosure: Only essential results are dis- played upfront, with detailed forensic evidence or raw model outputs available on demand.
- Simplified visualization: Graphical representations such as bar charts, line trends, or anomaly scores are preferred over dense numerical tables.

D. Multimodal Detection with Feedback

To improve robustness and user engagement, the framework integrates multiple data modalities and iterative feedback:

- Audio-visual fusion: Combines visual artifacts, facial ex- pressions, and audio cues to enhance detection accuracy in diverse real-world scenarios.
- Temporal consistency checks: Detects inconsistencies across consecutive frames or audio segments, flagging sudden anomalies.
- User feedback loop: Users can confirm or challenge detection results, which the system logs to refine future predictions and improve explainability.
- Adaptive thresholds: Detection sensitivity can be ad- justed based on context or user preference, balancing false positives and false negatives.

E. Ethical and Privacy Considerations

The framework also incorporates ethical safeguards:

- Data privacy: User-uploaded media is processed se- curely, with no retention of personal data without consent.
- Bias mitigation: Model outputs are monitored to reduce demographic or modality-specific biases.
- Transparent reporting: Users are informed about sys- tem limitations, including false-positive rates and scenar- ios where detection is less reliable.

This framework emphasizes a **human-in-the-loop ap- proach**, combining technical accuracy with interface design, transparency, and ethical safeguards, thereby improving real- world adoption of deepfake detection tools.

TABLE I
COMPARISON OF DEEPFAKE DETECTION APPROACHES

Method	Modality	Accuracy	Explainability
Visual Artifacts	Video	High	Low
Frequency Analysis	Video	Medium	Low
Multimodal Detection	Audio+Video	High	Medium
HCI-Centered	Multimodal	High	High

Approach

IV. COMPARATIVE ANALYSIS

Table I presents a comparative analysis of widely used deepfake detection approaches based on detection modal- ity, accuracy, and explainability. Traditional visual artifact- based and frequency analysis techniques demonstrate high to moderate accuracy in controlled environments; however, they offer limited interpretability, making it difficult for users to understand detection outcomes. Multimodal detection approaches, which combine audio and video cues, improve robustness and provide a moderate level of transparency. In contrast, the proposed HCI-centered approach integrates multimodal detection with user-centered interface design and explainable AI techniques. This inte- gration enables high detection accuracy while significantly enhancing system interpretability and user trust, making it more suitable for real-world deployment.

V. RESULTS AND ANALYSIS

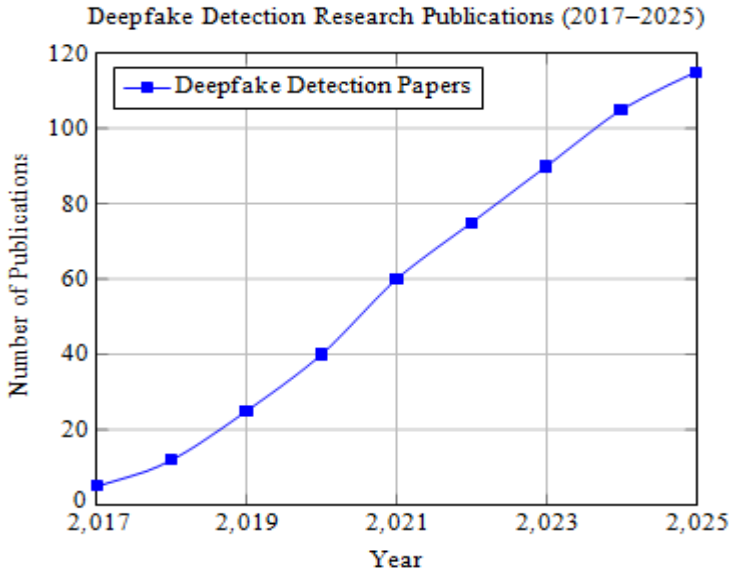


Fig. 1. Year-wise growth of deepfake detection research publications from 2017 to 2025. illustrates the rapid growth of deepfake detection research. Over 500+ papers have been published since 2018, reflecting the increasing atten- tion to this domain. Systematic reviews indicate that, despite this proliferation, detection models often experience performance decay in real-world scenarios, highlighting the importance of user-centered and explainable approaches.

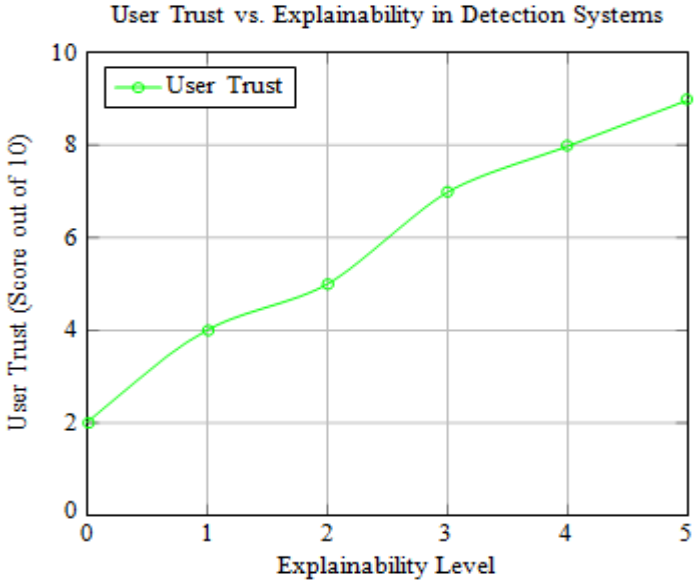


Fig. 2. Relationship between user trust and explainability in deepfake detec- tion systems. Empirical findings indicate that user

trust increases significantly when detection systems provide explanations alongside predictions.

VI. CHALLENGES

Deepfake detection systems face persistent challenges, in- cluding:

1) Performance decay on unseen data.
2) Vulnerability to adversarial attacks.
3) Trade-offs between accuracy and usability.
4) Ethical concerns such as labeling, censorship, and user autonomy.
5) Poorly designed interfaces leading to user skepticism or cognitive overload.

Addressing these challenges requires combining **technical performance with HCI-centered design principles**, ensuring systems are both accurate and user-friendly.

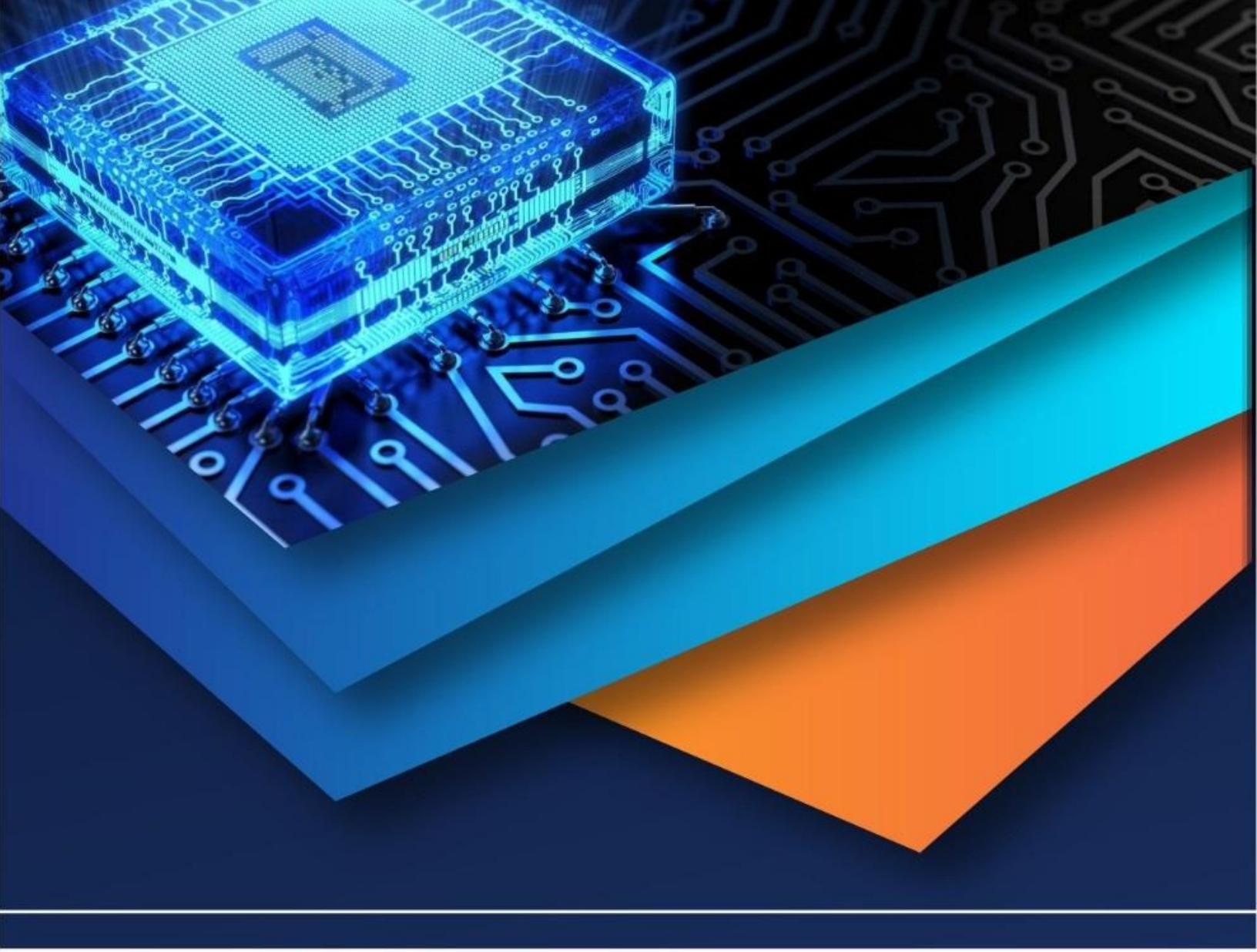
VII. CONCLUSION

This paper emphasizes that deepfake detection is not solely a technical problem but a socio-technical challenge requiring human-centered solutions. Integrating HCI principles into detection systems improves usability, transparency, and trust, which are critical for real-world adoption. This study demonstrates that effective deepfake detection extends be- yond algorithmic accuracy and must be approached as a human–computer interaction problem.

By combining explainable AI, cognitive load-aware inter- face design, and multimodal detection, systems can better support users in navigating synthetic media. Future work should explore adaptive learning strategies and cross-cultural usability studies to address evolving deepfake threats. Aligning algorithmic performance with human needs is essential for building trustworthy and ethically responsible detection systems.

REFERENCES

[1] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “MesoNet: A compact facial video forgery detection network,” in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), 2018.
[2] X. Yang, Y. Li, and S. Lyu, “Exposing deepfakes using inconsistent head poses,” in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2019.
[3] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, “Protecting world leaders against deepfakes,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.
[4] P. Korshunov and S. Marcel, “Deepfakes: A new threat to face recog- nition? Assessment and detection,” arXiv preprint arXiv:1904.07399, 2019.
[5] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega- Garcia, “Deepfakes and beyond: A survey of face manipulation and detection,” Information Fusion, vol. 64, pp. 131–148, 2020.
[6] Y. Mirsky and W. Lee, “The creation and detection of deepfakes: A survey,” ACM Computing Surveys, vol. 54, no. 1, pp. 1–41, 2021.
[7] H. Khalid, S. Woo, and S. Lee, “FakeAVCeleb: A novel audio-video multimodal deepfake dataset,” in Proc. Int. Conf. on Multimedia Re- trieval (ICMR), 2021.
[8] G. Gupta, K. Raja, M. Gupta, T. Jan, S. T. Whiteside, and M. Prasad, “A comprehensive review of deepfake detection using advanced machine learning and fusion methods,” Electronics, vol. 13, no. 1, p. 95, 2024.
[9] T. P. Nagarhalli, A. Save, S. Patil, and U. Aswalekar, “A comprehensive review of deepfake and its detection techniques,” SSRG International Journal of Electrical and Electronics Engineering, 2024.
[10] K. Somoray, D. Miller, and M. Holmes, “Human performance in deep- fake detection: A systematic review,” Human Behavior and Emerging Technologies, 2025.
[11] J. Richings, M. Leblanc, I. Groves, and V. Nockles, “Performance decay in deepfake detection: The limitations of training on outdated data,” arXiv preprint arXiv:2511.07009, 2025.
[12] A. E. Smith and L. Chen, “Explainable AI and user trust in human- centered security systems,” IEEE Computer, vol. 56, no. 3, pp. 45–53, 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY