

Detecting Ai Agents In Online Meetings Using Cybersecurity Techniques

A Multi-Model Cyber Security Framework for Impersonation Detection

¹Mr.G.Ravinder, ² Goparoju Neha, ³ Avula Srividhya, ⁴ Annam Manoj, ⁵ Sakinala Chandrakanth

¹Assistant Professor, ²³⁴⁵ Student

¹²³⁴⁵Department of CSE(Cyber Security)

¹²³⁴⁵ Siddhartha Institute of Technology & Sciences, Hyderabad, India.

Abstract—The rapid adoption of online meeting platforms has introduced new cybersecurity threats in the form of AI-driven impersonation and automated agents. Advanced generative models can now synthesize realistic human voices, facial expressions, and conversational behavior, enabling unauthorized AI agents to infiltrate virtual meetings and perform identity spoofing and social engineering attacks. This paper proposes a cybersecurity-oriented framework for detecting AI agents in online meetings by integrating multi-modal analysis techniques. The proposed system combines voice forensics, visual spoofing detection, behavioral biometrics, and anomaly-based intrusion detection to continuously verify participant authenticity. Rather than relying solely on traditional login-based authentication, the framework adopts a Zero Trust approach, where each participant is dynamically assessed throughout the session. Evaluation is conducted using benchmark datasets and simulated meeting scenarios to demonstrate the feasibility and effectiveness of the proposed approach. The results indicate that multi-layer cybersecurity analysis can significantly enhance trust, privacy, and security in online communication environments.

Keywords: AI impersonation, Deepfake detection, Zero Trust, Behavioral biometrics, Cybersecurity, Online meeting security.

I. INTRODUCTION

Online meeting platforms have become a primary medium for communication in corporate, educational, and governmental environments. While these platforms enable efficient remote collaboration, they also introduce significant cybersecurity challenges related to identity verification and unauthorized access. Recent advances in artificial intelligence have intensified these challenges by enabling AI-driven agents to convincingly mimic human voices, facial expressions, and conversational behavior. AI-based impersonation poses a serious security threat, as adversaries can use synthetic voice models, deepfake video generation, and large language models to infiltrate meetings and conduct identity spoofing or social engineering attacks. Traditional authentication mechanisms such as passwords, meeting links, and one-time verification are inadequate against such threats, as they do not provide continuous assurance of participant authenticity after initial authentication. From a cybersecurity perspective, AI-generated participants represent a form of identity spoofing and insider threat operating within trusted communication environments. To address this issue, this paper proposes a cybersecurity-oriented framework for detecting AI agents in online meetings. The framework integrates voice forensics, visual spoofing detection, behavioral biometrics, and anomaly-based intrusion detection, following principles inspired by Zero Trust security models. By applying layered cybersecurity analysis, the proposed approach aims to enhance trust and security in virtual meeting environments.

II. EXISTING SYSTEM

Current online meeting platforms primarily rely on conventional authentication mechanisms such as usernames, passwords, meeting links, and one-time passcodes to verify participant identity. While these methods are effective for initial access control, they are insufficient for defending against AI-driven

impersonation attacks once a session has begun. After authentication, participants are generally trusted for the entire duration of the meeting, creating security gaps that can be exploited by malicious AI agents.

Existing detection approaches for AI-generated content are mostly single-modal and limited in scope. Voice spoofing detection techniques analyze speech characteristics but often fail when AI-generated voices are enhanced using noise reduction or voice conversion models. Similarly, face spoofing and deepfake detection methods primarily focus on static images or offline video analysis, making them less effective in dynamic online meeting environments. CAPTCHA-based verification and manual monitoring by meeting hosts are also commonly used, but these methods are disruptive and unreliable for continuous identity assurance.

From a cybersecurity standpoint, current systems lack continuous verification, behavioral monitoring, and intrusion detection capabilities. They do not treat AI-driven participants as potential insider threats and fail to integrate multi-layer security analysis. As a result, existing systems are unable to reliably detect AI agents that exploit human-like behaviour to bypass traditional access controls in online meetings.

III. PROPOSED SYSTEM

The proposed system presents a cybersecurity-focused framework for detecting AI agents in online meetings through continuous participant assessment. Unlike traditional approaches that rely only on initial authentication, the framework follows principles inspired by Zero Trust security models, where participant trust is continuously evaluated throughout the session.

The system employs a multi-layer analysis approach by examining audio, video, and behavioural characteristics. Voice forensics techniques are used to identify synthetic speech artifacts, while visual spoofing detection analyzes facial inconsistencies associated with deepfake content. Behavioural biometrics further support detection by monitoring interaction patterns and response timing for anomalies. Outputs from these modules are integrated using an anomaly-based intrusion detection and risk scoring mechanism that estimates the likelihood of AI-driven impersonation. By combining multi-modal analysis with cybersecurity risk assessment, the proposed framework enhances protection against identity spoofing and AI-based infiltration in online meeting environments.

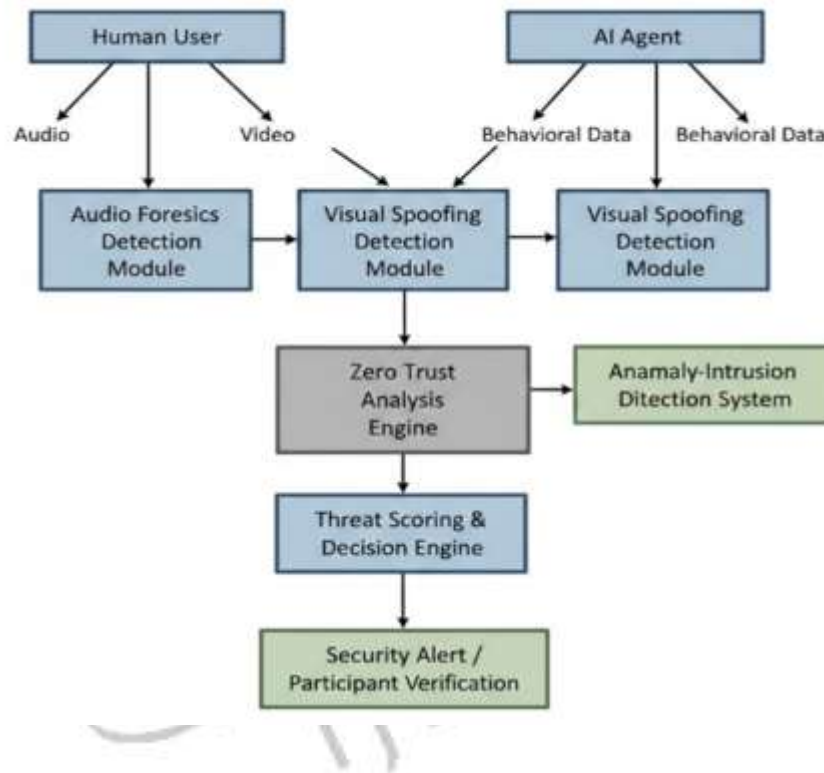
SYSTEM ARCHITECTURE

The implementation of the proposed framework is designed in a modular manner to support cybersecurity-driven analysis of online meeting participants. The system is structured as independent modules that process audio, video, and behavioural data streams extracted from meeting sessions or benchmark datasets.

Audio data is processed using voice forensics techniques, where features such as Mel-frequency cepstral coefficients (MFCCs) and spectral patterns are extracted to identify characteristics commonly associated with AI-generated speech. Video data is analyzed by extracting facial frames and examining visual inconsistencies indicative of deepfake manipulation. Behavioural data, including response timing and interaction patterns, is analyzed to detect anomalies that may suggest automated or non-human behaviour.

Each module produces an independent risk indicator, which is then aggregated by an anomaly-based intrusion detection and scoring engine. This engine evaluates deviations from expected human behaviour and assigns a composite risk score representing the likelihood of AI-driven impersonation.

The modular design allows the framework to be evaluated using benchmark datasets and simulated meeting scenarios, and it can be extended or adapted for integration with online meeting platforms in future deployments.



IV. IMPLEMENTATION

The implementation of the proposed framework is structured in a modular manner to support cybersecurity-based detection of AI-driven impersonation in online meetings. The system processes audio, video, and behavioral data streams independently, enabling flexible evaluation and future extensibility.

Audio data is analyzed using a voice forensics module that extracts speech features such as Mel-Frequency Cepstral Coefficients (MFCCs) and spectral representations. These features are examined to identify artifacts commonly introduced by AI-generated voice synthesis.

Video data is processed through a visual spoofing detection module, where facial frames are extracted and analyzed for spatial and temporal inconsistencies associated with deepfake manipulation. Behavioral data is analyzed by monitoring interaction patterns such as response timing and conversational flow to identify deviations from normal human behavior.

The outputs generated by each module are forwarded to a centralized threat scoring engine, which aggregates the results and produces a final participant verification decision. The modular design enables the framework to be evaluated using simulated inputs and benchmark datasets.

V. ALGORITHMS

The proposed framework employs multiple algorithms designed to detect AI-driven impersonation by analyzing audio, visual, and behavioural characteristics. Each algorithm functions as a security detection module, contributing to the overall intrusion detection process.

Audio Forensics-Based Spoof Detection Algorithm:

This algorithm extracts speech features such as MFCCs and spectral representations from participant audio. These features are analyzed to identify artifacts commonly introduced by AI-generated voice synthesis. The output is a probability score indicating the likelihood of synthetic speech.

Visual Spoofing Detection Algorithm:

Facial frames are extracted from video streams and analyzed for spatial and temporal inconsistencies associated with deepfake manipulation. A convolutional neural network-based classifier evaluates frame-level anomalies, and a voting mechanism aggregates the results to determine the likelihood of visual spoofing.

Behavioural Biometrics Algorithm:

This algorithm monitors interaction patterns such as response timing and conversational flow. Deviations from expected human behavioural patterns are treated as anomalies and used to estimate the probability of automated or AI-generated behaviour.

Threat Scoring and Decision Algorithm:

The outputs of all detection modules are fused using an anomaly-based risk assessment mechanism. A composite threat score is computed to represent the overall likelihood of AI-driven impersonation, enabling the system to flag suspicious participants for further verification.

OUTPUT

```
==== FINAL THREAT ANALYSIS ====
```

```
Audio AI Probability : 0.78
```

```
Video AI Probability : 0.72
```

```
Behavior AI Probability: 0.69
```

```
AI-Likeness Risk Score : 0.73
```

```
⚠️ ALERT: AI Agent Detected
```

```
==== FINAL THREAT ANALYSIS ====
```

```
Audio AI Probability : 0.21
```

```
Video AI Probability : 0.18
```

```
Behavior AI Probability: 0.25
```

```
AI-Likeness Risk Score : 0.21
```

```
✅ Participant Verified as Human
```

VII . CONCLUSION

The increasing use of AI-generated voices, videos, and automated agents in online communication platforms has introduced serious cybersecurity risks related to identity spoofing and social engineering. This paper presented a cybersecurity-oriented framework for detecting AI agents in online meetings by integrating voice forensics, visual spoofing detection, behavioural biometrics, and anomaly-based intrusion detection techniques.

By adopting a layered security approach inspired by Zero Trust principles, the proposed framework emphasizes continuous assessment of participant authenticity rather than relying solely on initial authentication. The study demonstrates the feasibility of applying multi-modal cybersecurity analysis to mitigate AI-driven impersonation threats in virtual meeting environments. Future work may focus on large-scale deployment, real-world evaluation, and improving robustness against adaptive AI-based attacks.

ACKNOWLEDGEMENT

We express our sincere gratitude to our project guide, faculty members, and department for their support, valuable suggestions, and encouragement throughout the project. We also thank our peers for providing feedback during the development and testing phases.

References

1. Muhammad Usama, Junaid Qadir, and Ala Al-Fuqaha, "AI in Cybersecurity: A Survey," *IEEE Access*, vol. 7, pp. 13394–13420, 2019.
2. Nicholas Carlini et al., "Adversarial Machine Learning in Real-World Systems," *ACM Computing Surveys*, vol. 55, no. 3, pp. 1–38, 2021.
3. Andreas Houmansadr, "Artificial Intelligence for Cybersecurity," *ACM Computing Surveys*, vol. 54, no. 5, pp. 1–36, 2021.
4. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., "FaceForensics++: Learning to Detect Manipulated Facial Images," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–11, 2019.
5. ASVspooof Consortium, "ASVspooof: Automatic Speaker Verification Spoofing and Countermeasures Challenge," 2021. [Online]. Available: Benchmark dataset for voice spoofing detection.
6. Wang, L., Liu, Y., and Zhang, Y., "Cybersecurity Threat Detection Using Machine Learning," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2346–2358, 2020.
7. Zampunieris, D., and Smith, J., "Machine Learning for Automated Incident Response," *IEEE International Conference on Intelligent Robots and Systems (IROS)*, pp. 432–438, 2018.
8. Sadasivam, S., "Real-Time Deepfake Detection Systems," *International Conference on Intelligent Computing and Communication Systems (ICICCS)*, pp. 456–461, 2020.
9. Zhang, Y., and Li, X., "Deepfake-Based Identity Attacks and Detection Methods," *Cybersecurity and Cyberwar (CyberC)*, pp. 89–94, 2021.
10. Zhu, H., Chen, M., and Wang, J., "Cyber Threat Detection Using Artificial Intelligence Techniques," *IEEE Access*, vol. 8, pp. 184512–184524, 2020.