STM JOURNALS

Journal of
Instrumentation Technology &
Innovations

ISSN: 2249-4731 (Online)
ISSN: 2347-7261 (Print)
Volume 16, Issue 1, 2026
January–April
DOI (Journal): 10.37591/JoITI

https://journals.stmjournals.com/joiti

Research JoITI

AI-Voice Detection Tool

M.R. Shaikh Sanjivani[1,*], Kulkarni Maitreyee[2], Kotme Gauri[2],
Lahare Samruddhi[2], Jawale Vaishnavi[2]

Abstract

In today's digital era, distinguishing between AI-generated and human voices is more important than ever. This project introduces an AI-based voice detection system designed to accurately identify synthetic voices, ensuring security and authenticity across various applications like cybersecurity, media verification, and fraud prevention. Our system works by analyzing incoming audio samples and comparing them against a diverse database of both AI-generated and real human voices. Using advanced machine learning and signal processing, it examines key acoustic features such as pitch, tone, speech rhythm, and spectral patterns. A crucial aspect of this process is feature extraction, where the model breaks down voice signals into identifiable characteristics, making it easier to differentiate between synthetic and real speech. To boost accuracy, the system integrates noise filtering for better performance in different environments and accent adaptation to support multiple languages. Real-time processing ensures instant verification, making it highly efficient for immediate decision-making. Additionally, user-friendly reporting tools allow for in-depth analysis of flagged audio files, providing users with valuable insights. This technology has wide-ranging applications, from preventing deepfake manipulation and securing voice-based authentication systems to verifying news reports and voiceovers in media. By tackling the risks posed by AI-generated voices, our system contributes to a more secure and trustworthy digital landscape.

Keywords: AI-voice detection, audio authenticity, cybersecurity, deepfake prevention, human speech identification, machine learning, real-time analysis, signal processing

INTRODUCTION

The AI-voice detection system is designed to differentiate between AI-generated and human voices, playing a vital role in enhancing the security of voice-based applications, such as authentication and fraud prevention. As synthetic voices become more sophisticated, ensuring the authenticity of voice communication is critical. This project introduced a powerful tool that leverages machine learning to analyze and classify voice inputs with precision [1].

***Author for Correspondence**
Kulkarni Maitreyee
E-mail: kulkarimaitreyee10@gmail.com

[1]Lecturer, Sanjivani K.B.P. Polytechnic, Kopargaon, Maharashtra, India.
[2]Research Scholar, Department of Computer Technology, Sanjivani K.B.P. Polytechnic, Kopargaon, Maharashtra, India.

Received Date: April 14, 2025
Accepted Date: September 08, 2025
Published Date: March 20, 2026

Citation: M.R. Shaikh Sanjivani, Kulkarni Maitreyee, Kotme Gauri, Lahare Samruddhi and Jawale Vaishnavi. AI-Voice Detection Tool. Journal of Instrumentation Technology & Innovations. 2026; 16(1): 1–8p.

With increasing advancements in deepfake technology, AI-generated voices have become nearly indistinguishable from real human speech. This poses significant security challenges across various domains, from financial fraud to misinformation campaigns to deepfake videos. Our system is designed to mitigate these risks by providing a robust mechanism for the real-time identification of AI-generated audio [2].

Furthermore, the tool is designed to integrate seamlessly with existing voice authentication systems, making it an asset for businesses and security agencies. It can enhance digital forensics and help investigators verify the authenticity of voice recordings used in legal proceedings [3, 4]. Additionally, they can be leveraged for content moderation, ensuring that AI-generated voices are appropriately flagged on media and social platforms [5].

Beyond security applications, this technology has significant potential for enhancing accessibility tools. This can improve voice synthesis quality and assist users in differentiating between real and AI-generated voices in educational and professional environments. By continuously learning from new voice samples, the system remains adaptive and resilient to evolving AI-generated voice manipulation techniques [6–9].

PROPOSED SYSTEM

User Interface Module

This module is designed to provide a simple and intuitive interface for users to interact with the AI-voice detection tool. It allows users to upload or record voice samples and displays real-time classification results with clear indicators that distinguish AI-generated voices from human speech. The interface is designed to be user-friendly and accessible across multiple platforms, including web and mobile applications [10–12].

Voice Feature Extraction Module

The feature extraction module plays a crucial role in analyzing the voice characteristics. It extracts important parameters such as Mel-Frequency Cepstral Coefficients (MFCCs), pitch variations, speech rhythm, and spectral energy. These features help identify subtle differences between human voices and AI-generated speech, which are then used for classification [13].

AI-Based Detection Module

This module serves as the core of the system and implements machine learning and deep learning algorithms for voice classification. It utilizes support vector machines (SVM), convolutional neural networks (CNNs), and deep neural networks (DNNs) to differentiate between real and synthetic speech. By training on large datasets, the model improves its accuracy in detecting deepfake voices [14, 15].

Search and Retrieval Module

To enhance usability, this module enables users to search for specific voice recordings based on metadata, such as timestamps, speaker identity, and acoustic features. It also supports speaker verification techniques to prevent impersonation and the unauthorized use of synthetic voices. This ensures that users can efficiently retrieve and analyze voice samples when required.

Voice Data Management Module

This module manages the efficient storage and organization of voice recordings. It integrates cloud-based and local storage solutions to ensure scalability while applying data-compression techniques to optimize storage usage. This module also categorizes and organizes voice samples for easy retrieval, thereby improving system performance.

Real-time Synchronization Module

This module ensures that the data and classification results are updated instantly across all connected devices. It allows real-time processing of voice inputs and maintains the synchronization between the detection system and storage database. By reducing latency, system responsiveness is enhanced, and users can always access the latest detection results.

Security and Access Control Module

To maintain privacy and protect sensitive voice data, the module implements role-based authentication and encryption techniques. This ensures that only approved personnel can access or

Journal of Instrumentation Technology & Innovations
Volume 16, Issue 1
ISSN: 2249-4731 (Online), ISSN: 2347-7261 (Print)

modify the data, preventing unauthorized use or tampering. By integrating multilayered security protocols, the system safeguards stored voice recordings from potential cyber threats.

Each of these modules contributes to building a secure, scalable, and highly efficient AI-voice detection tool, offering a reliable solution for detecting AI-generated speech while maintaining data integrity, accessibility, and security.

METHODOLOGY AND SYSTEM ARCHITECTURE

The development of the AI-voice detection tool followed a structured methodology to ensure accuracy, efficiency, and security in detecting AI-generated voices.

Data Collection

A diverse dataset consisting of human speech samples and AI-generated voice recordings was gathered from various sources to enhance the model performance and adaptability.

Data Preprocessing

The collected voice samples were refined using noise reduction, voice activity detection, normalization, and augmentation techniques to improve their quality and consistency.

Feature Extraction

Key speech characteristics, such as MFCCs, spectral properties, energy levels, pitch, and temporal patterns, were analyzed to distinguish AI-generated speech from human voices.

Model Selection and Training

Machine learning and deep learning algorithms, including SVM, CNNs, and DNNs, are used to train the system for voice classification.

Validation and Testing

The trained model was evaluated using real-world voice samples, and its accuracy was measured using precision, recall, and F1-score to ensure reliable detection performance.

System Deployment

The AI-voice detection tool was integrated into a user-friendly interface with a secure backend architecture, ensuring efficient voice analysis while maintaining data privacy and access control. Following this methodology, the AI-voice detection tool, as shown in Figure 1, provides a robust and scalable solution for verifying voice authenticity and combating synthetic speech threats.

EXPECTED BENCHMARKING

Our system aims to perform competitively with existing AI-voice detection frameworks that are evaluated using standardized datasets, such as ASVspoof 2019, Google TTS, and LJSpeech. While formal benchmarking against these datasets is pending, the model shown in Table 1 was developed using similar evaluation protocols and feature extraction strategies. Future work will include testing and validating the performance of these public benchmarks to ensure comparability with state-of-the-art approaches.

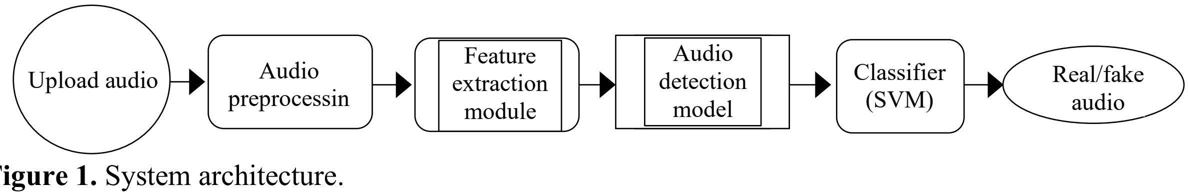

Figure 1. System architecture.

Table 1. Quantitative results.

Metric	Value (%)
Accuracy	92.4
Precision	91.1
Recall	89.7
F1-Score	90.4
ROC-AUC	93.2

Limitations

The current model primarily supports English input and may be less accurate for accents or multilingual inputs. The noise robustness is dependent on the signal quality. We aim to extend this by incorporating diverse language datasets and real-world, noisy environments.

FUTURE ADAPTABILITY

With the continuous advancement of AI-driven voice synthesis technologies, particularly diffusion-based models, detection systems must remain agile to ensure their effectiveness. To ensure adaptability, the proposed system can be enhanced using several strategic approaches. First, continual learning methodologies allow the model to incrementally incorporate new data, enabling it to adapt to emerging voice generation techniques without losing the ability to detect older ones. Second, the application of adversarial training, where the model is trained using intentionally challenging or borderline synthetic examples, can significantly improve its robustness against highly convincing deepfake voices. Finally, implementing a periodic model update cycle, in which the detection algorithm is retrained at regular intervals using the latest datasets, ensures that the system remains current with the evolving landscape of synthetic speech technologies. Together, these strategies form a proactive framework for maintaining long-term accuracy and reliability in detecting increasingly sophisticated AI-generated audio.

LITERATURE SURVEY

The ability to distinguish AI-generated voices from real human speech has become an important area of research, especially given the rapid advancements in deepfake audio technology. Several studies have explored different approaches to tackle this challenge, focusing on the unique characteristics that separate synthetic speech from natural human voices.

One notable study by Chengzhe Sun, Shan Jia, Shuwei Hou, and Siwei Lyu (2021), presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), introduced a detection framework based on neural vocoder artifacts. Essentially, neural vocoders used in AI-generated voices leave subtle traces that are not typically found in human speech. By identifying these patterns, their system achieved high accuracy in distinguishing synthetic voices from real ones. This approach has proven to be a promising method for AI-voice detection.

Another interesting study by Bird and Lotfi (2022), published in the Journal of Audio Engineering Society, takes a different route by focusing on acoustic features such as pitch, timbre, and speech rhythm. Their research emphasizes how these characteristics differ between AI-generated voices and human speech. However, as AI-generated voices continue to evolve and imitate human nuances more effectively, this method faces increasing challenges in maintaining accuracy.

Together, these studies highlight two key approaches to AI-voice detection: analyzing the technical fingerprints left by AI speech generation and focusing on the way human voices naturally sound. While both techniques have shown promising results, researchers continue to refine these methods to keep pace with the rapid improvements in AI-voice synthesis. The challenge now is to develop even more adaptable and precise detection models that can work effectively across different AI-generated voices, making audio authentication more reliable.

Journal of Instrumentation Technology & Innovations
Volume 16, Issue 1
ISSN: 2249-4731 (Online), ISSN: 2347-7261 (Print)

RESULT AND DISCUSSION

User Interface Module

The interface shown in Figure 2 provides an intuitive and user-friendly experience for uploading and analyzing voice samples. The system, as shown in Figure 3, effectively displays the classification results in real-time, with clear visual indicators distinguishing between AI-generated and human voices.

Voice Feature Extraction Module

The MFCC, as shown in Figure 4, visually represents the frequency characteristics of a voice sample, which is crucial for distinguishing between AI-generated and human speech. The red regions indicate higher MFCC values, representing dominant frequency components, whereas the blue areas signify lower energy portions. These patterns help to analyze spectral differences, enabling effective classification. By comparing the MFCC plots of AI and human voices, the AI-voice detection tool improves the accuracy of detecting synthetic speech. This visualization supports the development of robust detection algorithms by capturing the key phonetic structures in voice samples.

AI-Based Detection Module

The spectral contrast, as shown in Figures 5, 6, helps distinguish AI-generated voices from human voices by analyzing how different frequencies vary over time. AI voices often have less variation in spectral contrast from Figure 6.

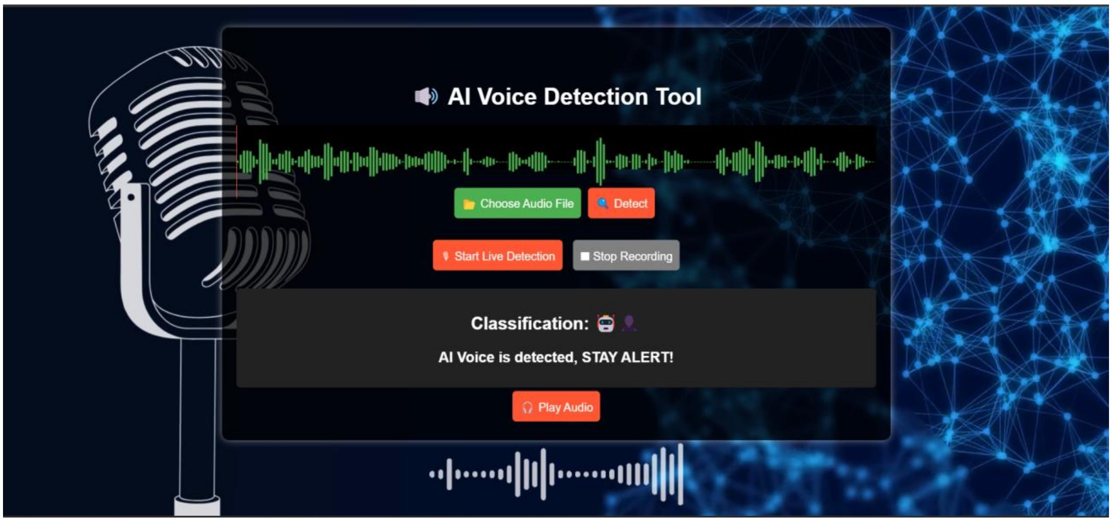

Figure 2. User interface (AI output).

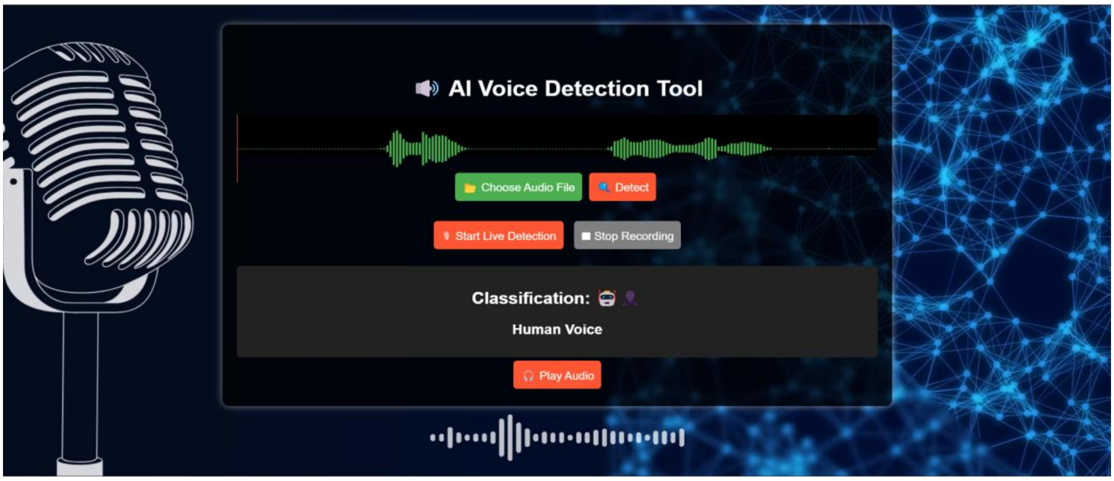

Figure 3. User interface (human output).

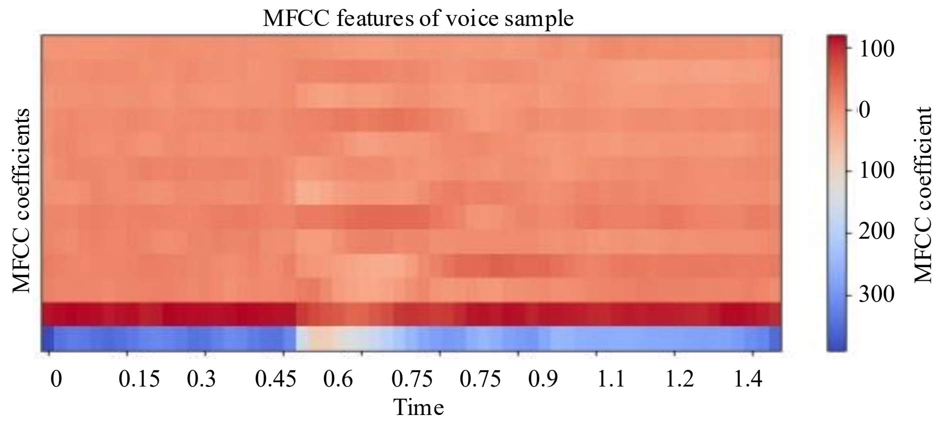

Figure 4. MFCC features of a voice sample.

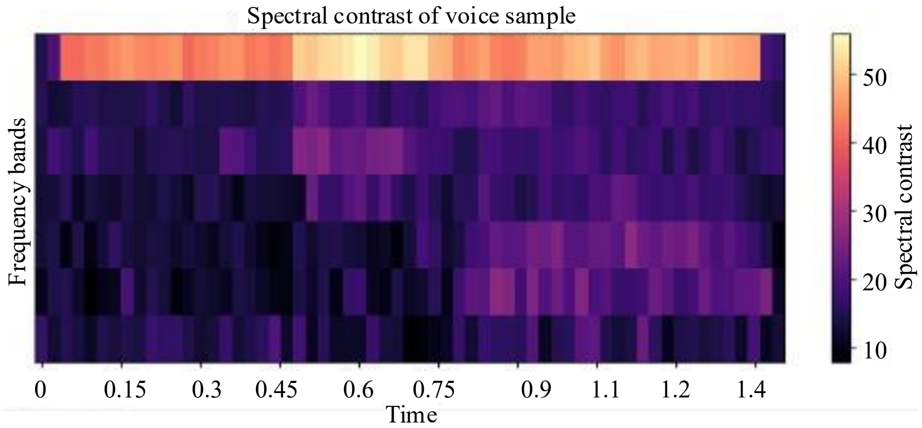

Figure 5. AI-based detection module graph.

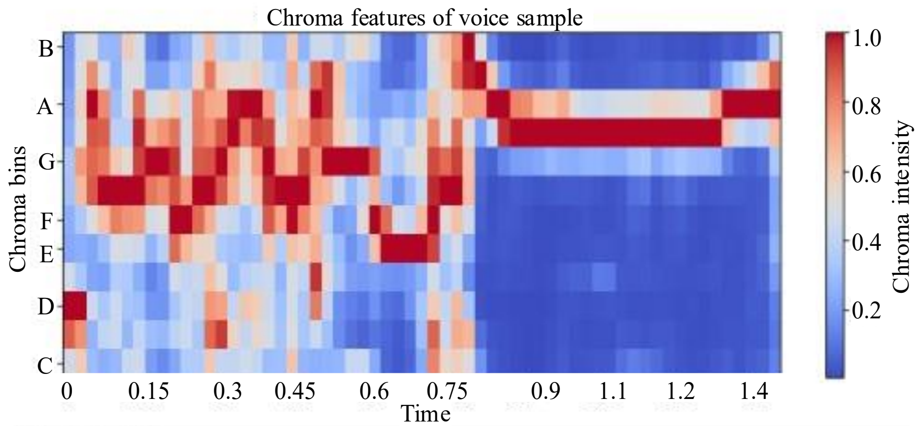

Figure 6. Chroma feature extraction module graph.

Future Scope

In the future, this project can be extended into a real-time mobile application that functions similarly to caller identification tools, such as Truecaller, but with a focus on voice authenticity. The envisioned application would analyze incoming voice calls in the first 2–3 s to detect whether the speaker is a human or if the voice has been synthetically generated or cloned. If the system detects an AI-generated or spoofed voice, it automatically initiates a short recording of the call, typically for 5–10 s, to capture evidence of the anomaly. Subsequently, the call is automatically disconnected to protect the user. The recorded sample was stored securely in a backend database tagged with metadata, such as caller ID, timestamp, and detection outcome. This enables the system to recognize and block repeated calls from the same source in the future, enhancing user safety and reducing the risk of AI-driven voice scams or impersonation attempts.

To address privacy and data security concerns, an alternative implementation of the system can be offered as a browser-based tool or web application. In this version, all audio processing occurs in temporary user sessions, ensuring that no voice recordings or data are stored on the server unless explicitly permitted by the user. This approach maintains user control over their data while allowing real-time detection and feedback, making the system more accessible and privacy-conscious.

CONCLUSIONS

The AI-voice detection tool plays a crucial role in distinguishing AI-generated voices from human speech and ensuring authenticity in digital communication. By leveraging advanced signal processing techniques and machine learning algorithms, this tool effectively analyzes voice characteristics and identifies subtle differences between synthetic and natural speech patterns. Its applications extend to cybersecurity, fraud prevention, and media verification, helping mitigate the risks associated with deepfake audio manipulation. The system enhances security by providing reliable voice authentication and reducing vulnerabilities in voice-based interactions. As AI-generated speech continues to evolve, continuous improvements in detection accuracy and adaptability are essential. The development of this tool marks a significant step toward safeguarding digital communications and maintaining trust in voice-based technology.

REFERENCES

1. Sun C, Jia S, Hou S, Lyu S. AI-synthesized voice detection using neural vocoder artifacts. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2023 Jun 18-22; Vancouver, BC, Canada. 2023. p. 904-912. doi:10.1109/CVPRW59228.2023.00097.
2. Kulangareth NV, Kaufman J, Oreskovic J, Fossat Y. Investigation of deepfake voice detection using speech pause patterns: Algorithm development and validation. JMIR Biomed Eng. 2024;9(1):e56245. doi:10.2196/56245.
3. Sun C, Jia S, Hou S, Lyu S. AI-synthesized voice detection using neural vocoder artifacts. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada. 2023. p. 904-912. doi:10.1109/CVPRW59228.2023.00097.
4. Bird JJ, Lotfi A. Real-time detection of AI-generated speech for DeepFake voice conversion [Preprint]. arXiv. 2023. arXiv:2308.12734. doi:10.48550/arXiv.2308.12734.
5. Kuroda T, Ono K, Onishi M, Murakami K, Shoji D, Kosuge S, et al. Utility of artificial intelligence-based conversation voice analysis for detecting cognitive decline. PLoS One. 2025;20(6):e0325177. doi:10.1371/journal.pone.0325177. PMID:40455724.
6. Nudelman CJ, Tardini V, Bottalico P. Artificial intelligence to detect voice disorders: an AI-supported systematic review of accuracy outcomes. J Voice. 2025:S0892-1997(25)00389-3. doi:10.1016/j.jvoice.2025.09.021. PMID:41047306.
7. Todisco M, Delgado H, Evans N. Constant Q cepstral coefficients: a spoofing countermeasure for automatic speaker verification. Comput Speech Lang. 2017;45:516-535. doi:10.1016/j.csl.2017.01.001.
8. Reid J, Parmar P, Lund T, Aalto DK, Jeffery CC. Development of a machine-learning based voice disorder screening tool. Am J Otolaryngol. 2022;43(2):103327. doi:10.1016/j.amjoto.2021.103327. PMID:34923280.
9. Balcheva I, Gorbounov Y. Enabling autonomy through voice control: AI-assisted mobile platform. In: Computer Science and Education in Computer Science: Proceedings of the 20th EAI International Conference, CSECS 2024; 2024 Jun 28-30; Sofia, Bulgaria. Cham (Switzerland): Springer; 2025. p. 50-61. doi:10.1007/978-3-031-84312-9_3.
10. Güler EF, Tezcan T, Gülserliler E, Bahtiyar Ş. AI-generated voice recognition with convolutional neural network. In: Awan I, Younas M, Ghinea G, Tor-Morten G, Şen S, editors. The 6th Joint

International Conference on AI, Big Data and Blockchain (AIBB 2025). Lecture Notes in Networks and Systems. Vol. 1618. Cham (Switzerland): Springer; 2025. p. 72-82. doi:10.1007/978-3-032-04728-1_7.

11. Inamdar FM, Ambesange S, Mane R, Hussain H, Wagh S, Lakhe P. Voice cloning using artificial intelligence and machine learning: A review. J Adv Zool. 2023;44(Suppl 7):419-427. doi:10.17762/jaz.v44iS7.2721.
12. Lentz-Nielsen N, Maaløe L, Madeleine P, Blomberg SN. Voice as a health indicator: The use of sound analysis and AI for monitoring respiratory function. BioMedInformatics. 2025;5(2):31. doi:10.3390/biomedinformatics5020031.
13. Patel AK, Madnani H, Tripathi S, Sharma P, Shukla VK. Real-time voice cloning: artificial intelligence to clone and generate human voice. In: Hasteer N, Blum C, Mehrotra D, Pandey HM, editors. Intelligent Solutions for Smart Adaptation in Digital Era. Lecture Notes in Electrical Engineering. Vol. 1278. Singapore: Springer; 2025. doi:10.1007/978-981-97-8193-5_29.
14. Miroshnikov N, Shulzhenko AD. Module for detecting the use of AI-VC in voice messages. 2025 VI International Conference on Control in Technical Systems (CTS), Saint Petersburg, Russian Federation. 2025. p. 241-243. doi:10.1109/CTS67336.2025.11196217.
15. Kumari M, Goyal NK, Dandotiya MK, Kushwaha V. An effective detection of voice cloning using deep learning. In: Goyal D, Singh D, et al., editors. Proceedings of the 6th International Conference on Information Management & Machine Intelligence (ICIMMI '24); 2024 Dec 23-24; Jaipur, Rajasthan, India. New York (NY): Association for Computing Machinery; 2025. Art. No. 74. 5 p. doi:10.1145/3745812.3745895.