


# A systematic literature review on AI voice cloning generator: A game-changer or a threat?

Genesis Gregorious Genelza<sup>1\*</sup> 

<sup>1</sup>College of Teacher Education & Junior High School Department, University of Mindanao Tagum College, Philippines. [genesis.genelza@umindanao.edu.ph](mailto:genesis.genelza@umindanao.edu.ph)

\*Corresponding author: [genesis.genelza@umindanao.edu.ph](mailto:genesis.genelza@umindanao.edu.ph)



**Abstract:** AI voice cloning, sometimes referred to as voice synthesis, uses machine learning to mimic the speech of an actual person. This system requires a certain quantity of voice data to assess and include each person's unique vocal characteristics. It can mimic the original voice after being trained to create similar sounds. This systematic literature review presents concrete information on whether AI Voice Cloning Generator is a game-changer or a threat. Collected findings showed that it positively impacts society as a game-changer in 21st-century innovation, but it can also threaten humanity without proper supervision and unethical use for scamming and fraud. AI voice cloning emerges as a potent game-changer with the potential to completely transform connections, entertainment, and interpersonal relationships. However, because technology may be used as a tool for deception and manipulation, it needs a well-rounded strategy that highlights its advantages while putting strict controls in place to deal with the moral and security risks currently facing humanity. Industry executives, legislators, and ethicists must collaborate to create these regulatory structures and guarantee that technological innovation is used to improve society while protecting against potential misuse or damage. These standards must prioritize authorization, confidentiality, and disclosure to ensure fair utilization.

**Keywords:** AI voice cloning technology, Artificial intelligence, Systematic Literature Review, Voice cloning, Voice synthesis

## 1. Introduction

AI voice cloning, sometimes referred to as voice synthesis, uses machine learning to mimic the speech of an actual person. This system requires a certain quantity of voice data to assess and include each person's unique vocal characteristics. It can mimic the original voice after being trained to create similar sounds. Deep learning techniques are commonly employed in constructing voice cloning models in the artificial intelligence (AI) industry. One common approach is to apply the recurrent neural network (RNN) model, specifically designed to handle sequential input, like voice. Computer-generated speech was the first application of online AI text-to-voice. An ancient text-to-speech technology converts text to artificial speech and allows votes to be used for computer-human communication. Online AI voice cloning software is the speech equivalent if you are familiar with the idea of deep video fakes. Developers can organize an audio data set and train an AI voice model to understand any text in the target voice with just a few minutes of recorded speech (Polaris Market Research & Consulting, 2023).

In addition, Arik, Chen, Peng, Ping and Zhou (2018) stated that voice cloning is a much-desired feature for individualized speech interfaces. It has been demonstrated that many speakers can produce high-quality

**Research Article:** This article is published by *Jozac Publishers* in the *Journal of Emerging Technologies (JET)*. This article is distributed under a Creative Common [Attribution International License \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/). **Conflict of Interest:** The author/s declared no conflict of interest.



speech using neural network-based voice synthesis. Adapting speakers involves honing a multi-speaker generative model using a small number of cloning samples. Speaker encoding is based on training a different model to infer a new speaker embedding from cloning audios and utilize it with a multi-speaker generative model. Even with a few cloning audios, both techniques can perform well regarding natural speech and likeness to the original speaker. Although speech adaptation can achieve more naturalness and similarity, the speaker encoding approach has much lower cloning time or memory requirements, which makes it suitable for low-resource deployment.

Artificial intelligence is used in the quickly emerging AI voice cloning technology field to produce synthetic voices that resemble real humans. This technology can completely transform many sectors and facets of our lives, including healthcare, education, and entertainment. AI voice cloning technology uses a dataset of audio recordings from a particular person to train a machine-learning model. The model gains the ability to recognize and imitate the individual's voice's distinctive qualities, including pitch, timbre, and accent. After training, the model may produce synthetic speech that closely resembles the training subject (Babu, 2023). However, in the contemporary workplace, applying theoretical knowledge to adjust to the shifting landscape of the professional environment is still a must, even though technology is everywhere and readily available (Genelza, 2022).

Furthermore, the process of creating artificial voices that resemble a target voice through speech synthesis is known as voice cloning. The exact replication of an agent through cloning makes it useful for voice assistants and real-time voice changers. The process usually needs a significant amount of recorded speech to construct a collection of datasets that can be used to develop a new voice model. In other words, voice clones are now possible by following a few easy procedures. Even though AI voice cloning is a fantastic technology, people's privacy rights might be violated, particularly when protecting personal data. Fraudulent actors can use voice recordings obtained legally to commit crimes such as fraud, blackmail, and impersonation (El Zarzour, 2023).

We have heard social media platforms like TikTok, Facebook, YouTube, etc., and the voices of celebrity singers like Taylor Swift, Ariana Grande, and more that identically sounded like they were singing songs from other celebrity singers. Some users of the platforms were amazed at how it works well in their versions, but some needed clarification as they had thought that the songs were demos or covers of the said artists. Even some celebrity singers and actors were alarmed by voice cloning as this generator can copy everything easily and be used to trick people with this. With all the information gathered in this article, this leaves us with a question: Is it a game-changer or a threat?

## **2. Literature review**

### **2.1. The pros and cons of AI voice cloning generator: The power it holds**

AI voice cloning has made considerable advancements in recent years. Moreover, it is now possible to use these computers to create highly realistic voices. A person can work as a marketer, game developer, or content creator in the AI voice cloning industry; it can help produce unique, personalized content that can be found in a rabble. In education, cloning the voices of historical figures provides contemporary occasions for interactive teaching and spirited storytelling. It can also be used in audiobooks. In the market for AI voice cloning, authors can read autobiographies in their voices, historical figures can narrate their tales in their voices, and celebrities can have their voices used to recount books. This results in captivating, excellent listening engagement. However, voice is a unique moderator and reliable biometric security method. Therefore, the perpetrators can use presentation bombardments, such as recorded speech, computer-altered voice, or synthetic voice, to trick voice biometric systems into believing they can hear a genuine, authorized user and grant access to meaningful data and narratives (Polaris Market Research & Consulting, 2023).

Also, the market for AI cloning is expanding due to the quick developments in AI technology, especially in machine learning (ML) and natural language processing (NLP). Thanks to these developments, more complex AI models that imitate human behaviors, speech patterns, and decision-making processes can be created. Creating virtual customer service representatives through AI cloning also makes it possible to address customer inquiries, offer round-the-clock support, and help with joint problems. This increases customer satisfaction, lowers costs, and improves customer service efficiency, encouraging the industry to adopt AI cloning technologies. Combining AI clones and AR and VR in games opens new player immersion and engagement possibilities. The gaming experience can be improved by interacting with AI characters

that display realistic movements, actions, and replies. AI clones can act as online rivals, allies, or mentors, enhancing gaming with their dynamic and customizable nature. AI clones that mimic real-world interactions and situations can improve AR and VR training simulations. AI clones, for instance, can function as virtual patients in the healthcare industry, allowing medical practitioners to practice diagnosing and treating a range of ailments in a lifelike virtual setting. This integration makes it easier to train repeatedly and safely, which enhances decision-making and skill-building (Grand View Research, 2023).

AI voice cloning can also be utilized in the entertainment industry to produce captivating and lifelike dubbing for multilingual films, video games, and other media. This may contribute to viewers' immersiveness and enjoyment of these entertainment formats. - Produce captivating and lifelike voiceovers for motion pictures, video games, and other media. AI voice cloning in education can be utilized to give pupils individualized learning experiences. AI may be used, for instance, to construct a vocal clone of a teacher that could instruct and provide feedback to students on an individual basis. AI voice cloning technology can potentially transform numerous other sectors and facets of our lives besides these particular uses. AI voice cloning, for instance, has the potential to foster new kinds of artistic expression, provide accessibility for those with impairments, and foster a more inclusive and connected global community. It is crucial to remember that voice cloning technology powered by AI has the potential to be abused. AI voice cloning, for instance, might produce phony audio recordings that people could use to con people or disseminate disinformation. To ensure that AI voice cloning technology is used for good and not harm, it is crucial to design ethical and responsible applications for it (Babu, 2023). Hence, instructors and the educational setting must also know students' strengths and weaknesses to comprehend an individualized learning area. Enhancing student control over what is taught in class and producing positive academic outcomes are the results of teaching the curriculum subject and being proficient in the language used for instruction. These factors increase the learner's exposure to and opportunity to understand the instruction material (Genelza, 2022; 2023).

An emerging topic in speech processing called voice cloning seeks to produce artificial speech that closely mimics the sounds of certain people. Recent years have seen a notable increase in interest in voice cloning, a field of study that is fast developing. The principal aim of this technology is to generate artificial speech that resembles the speech of a particular speaker—the so-called cloned speaker. This method has excellent promise across several industries, especially in the media sector. Long-form text reading, including emails and web pages, voice-overs, dubbing, audiobook narration, and more are among the applications (Dale, 2023; González-Docasal & Álvarez, 2023). To guarantee that the online course is managed well, it is advised that the outcome be used, enhanced, and put into practice (Batalla et al., 2023).

However, one of the main worries concerning AI voice cloning is how simple it is to copy and modify someone's voice. Access to this kind of data can lead to various fraudulent activities, such as disseminating fake information, voice phishing, and social engineering. Identity theft and voice-based scams represent a severe threat in a country as diverse and heavily populated as India, where the languages and cultural landscapes might differ substantially. Furthermore, the ethical aspects and implications of AI voice cloning cannot be ignored. It is against the law to use someone else's voice without that person's permission, and it also raises moral concerns around consent and ownership. Voice cloning is quickly emerging as a popular technique for AI scam calls. Online audio clips of people's voices are used by con artists, who upload them to specialist websites that can mimic the voices. Although these apps have been around for a while, they are now more widely available and reasonably priced thanks to developments in generative AI technology (El Zarzour, 2023).

Moreover, growing implications of generative AI in the speech domain allow for real-time voice conversion from one human to another and even voice cloning. Real-time speech recognition for Deepfake Voice Conversion is critical since this technology presents a severe ethical risk and may result in privacy violations and deception. The significance of generative artificial intelligence (AI) has been significantly increasing in recent years. Cutting-edge devices can instantly change a speaker's voice using a microphone and complex deep-learning models. Thanks to advancements in consumer-level computing technology, it is now feasible to clone someone's speech and utilize it in a phone call or online conversation. While this technology might be appealing for entertainment, it presents a serious security risk as it develops. Humans frequently go unchecked because they utilize speech to identify people in social contexts. Since voice conversion can be exploited unethically to compromise security and privacy, voice recognition is also used for biometric authentication. In social engineering attacks, AI-generated speech, such as impersonation, can

be used maliciously. This is why systems must offer transparency in real-time regarding the legitimacy of a human voice during a conference or phone call, especially given the rapidly improving quality of these systems. The contributions of this work include an original dataset for audio classification that is made available for use in future research, a thorough examination of the statistical significance of audio features taken from both natural and artificial intelligence-generated speech, and, lastly, the development of machine learning models that are capable of real-time legitimacy prediction (Bird & Lotfi, 2023).

Meanwhile, the advancement of voice cloning technology presents moral questions about consent and identity. It is crucial to get support because it is only sometimes apparent if people have granted their permission for their voices to be utilized. To guarantee that people have the right to decide how their voice is used, it is vital to establish precise criteria for gaining consent for voice cloning. Furthermore, when a synthetic voice is identical to a genuine voice, it might be utilized to harm others by impersonating them. Thus, laws must be implemented to stop the improper use of voice cloning technology. By considering these ethical issues, we can guarantee that the potential advantages of voice cloning technology are fulfilled while limiting the potential hazards and downsides (Espinosa, 2023). Hence, administrators should also plan meetings, training sessions, or other professional development activities. They were able to identify their areas of strength and weakness with the aid of the seminar sessions (Aguimlod et al., 2023).

Accepting this technology and creating morally and responsibly appropriate uses for it is imperative. AI voice cloning can provide lifelike and captivating content for customer support, education, healthcare, and entertainment. It can also be used to increase accessibility for individuals with impairments and to create a more inclusive and connected environment. However, it is crucial to be mindful of the possible drawbacks of AI voice cloning technology, such as the potential for abuse to disseminate false information or con people. We can reduce these concerns by producing accurate and exciting material, being open about AI voice cloning, and concentrating on applying the technology for the sake of society.

### 3. Methodology

This research article utilized a systematic literature review (SLR) as a qualitative methodology to address the AI Voice Cloning Generator issue. The systematic literature review is a methodical and purposeful way to locate, evaluate, and summarize previous research relevant to a particular subject or inquiry. A comprehensive search is carried out across several databases, academic journals, books, and other sources to find all pertinent research on the topic of interest. The systematic review's plan or methodology should be well stated, and the criteria should be fully specified before the study. This comprehensive and transparent search can be replicated by other researchers using many databases and grey literature sources. This involves planning a thorough search strategy and clearly emphasizing answering an identified question (Dewey & Drahota, 2016).

### 4. Discussion

Based on the findings drawn from those selected papers, the assertions that follow and this credible information should be emphasized in this article and shown in Table 1:

**Table 1:** List of Literature on AI Voice Cloning Generator

Authors	Title of the Study	Findings	Recommendations
Babu (2023)	Embracing AI voice cloning technology	Healthcare, education, and entertainment are just a few areas and aspects of our lives that AI voice cloning fundamentally affects. AI voice cloning method trains a machine-learning model with a dataset of audio recordings from a specific individual. Pitch, timbre, and accent are just a few characteristics of a human voice that the model learns to identify and mimic.	It must be remembered to keep in mind that AI-powered voice cloning technology has the potential to be exploited. For example, artificial intelligence voice cloning could result in fake audio recordings that people could use to deceive others or spread false information. Developing morally sound uses for AI voice cloning technology is essential to ensure it is used for good

			and not evil.
<b>Bird and Lotfi (2023)</b>	Real-time Detection of AI-Generated Speech for Deepfake Voice Conversion	Modern technology employs sophisticated deep-learning algorithms with a microphone to instantaneously alter a speaker's speech. Due to advances in consumer-level computing technologies, I can now clone someone's speech and use it in a phone call or online interaction.	While this technology may be entertaining, it poses a significant security concern as it advances. Because humans use speech to identify persons in social circumstances, they are often left uncontrolled. Voice recognition is also used for biometric authentication since voice conversion can be used unethically to jeopardize privacy and security.
<b>El Zarzour (2023)</b>	Caution Ahead: Exploring the Benefits and Dangers of AI Voice Cloning.	Voice assistants and real-time voice changers can benefit from the precise replication of a voice achieved by cloning.	AI voice cloning is a remarkable technology, but it may breach people's privacy rights, especially when protecting personal information. Voice recordings that were lawfully obtained can be used by dishonest actors to perpetrate crimes like impersonation, fraud, and blackmail.
<b>González-Docasal, and Álvarez (2023)</b>	Enhancing Voice Cloning Quality through Data Selection and Alignment-Based Metrics	This technology's primary objective is to produce synthetic speech that sounds like the speech of a specific speaker—a so-called "cloned speaker." This approach has great promise in several industries, particularly the media.	Be vigilant in using the technology, as this can also threaten humanity, not just a blessing to everyone.
<b>Polaris Market Research &amp; Consulting (2023)</b>	AI Voice Cloning Market: A Prospective Technique for Assimilating Distinct Vocal Features.	Authors can read autobiographies in their voices, historical figures can tell their tales in their agents, and celebrities can have their voices used to recount novels, thanks to the growing industry for AI voice cloning. As a result, there is exceptional and engaging listener engagement.	A dependable biometric security technique and distinct moderator is voice. To fool voice biometric systems into thinking they can hear a genuine, authorized user and provide access to essential data and narratives, the criminals can thus use presentation bombardments, such as recorded speech, computer-altered voice, or synthetic voice.

This review acquired two thematic points: a game-changer and a threat. Based on the articles presented in Table 1, AI voice cloning is a revolutionary innovation with several applications in different industries. It allows for a more customized and exciting user experience, first and foremost. Artificial intelligence (AI) voice cloning can improve accessibility, education, and enjoyment by imitating a particular voice, whether that of a historical person, celebrity, or loved one. It also makes it possible to create content without the original speaker being present, saving time and money in sectors like entertainment and advertising. Furthermore, AI voice cloning can provide a more natural expression for people with speech problems, promoting inclusivity and improving communication.

Nevertheless, besides its advantages, AI voice cloning raises moral and security questions. Consent and privacy are two critical issues. When someone's voice is used without permission to produce synthetic content, it can be misused to make deepfake movies or misleading information, harming the person's credibility. Furthermore, as this technology advances, there is a greater chance of identity theft and impersonation, which raises data security concerns. The production of phony audio that is incredibly lifelike has the potential to damage recorded evidence, which could affect the forensic and legal fields.

It is more important than ever to create thorough laws and regulations because of the possible risks and negative consequences that AI voice cloning may have. These should cover every privacy, ethical, and legal concern raised by voice cloning technology. One crucial component is strengthening the current data protection laws to include voice recordings as sensitive personal information. Strict measures must be taken to ensure that consent is granted and that voice data is securely preserved for a finite amount of time. AI voice cloning has gained a lot of attention and reputation recently. It involves creating artificial sounds that replicate genuine voices—actual humans. While there are benefits to this technology, there are also a lot of uncertainties and possible hazards. It is best to confront these problems as soon as feasible to deal with them proactively. We can reduce any potential adverse effects of AI voice cloning and provide future protections by addressing its issues as soon as they arise (El Zarzour, 2023).

Artificial intelligence (AI) voice cloning can potentially transform communication, but it also raises serious ethical and security issues. Harnessing innovation's promise while preventing potential misuse and preserving individual rights and privacy will require striking a balance between ethical considerations and innovation.

## **5. Recommendation**

AI voice cloning has much potential if used legally and wisely. It must be carefully integrated across multiple sectors to realize its potential fully. First off, AI voice cloning has the potential to revolutionize the healthcare industry by giving those with speech difficulties a more natural way to communicate. Furthermore, by providing more individualized and human-like interactions, AI voice cloning integrated into interactive voice response systems can improve the user experience in customer service. Increased customer satisfaction and more effective query management could result from this matter.

Further, an AI voice cloning generator is a tool that entertainment companies can use to produce engrossing and stimulating material. It can be used, for example, to bring historical individuals' voices back to life in educational materials or to personalize entertainment by letting consumers engage with their favorite characters in virtual reality and gaming environments. Companies could also investigate its possibilities in marketing and advertising by utilizing well-known or celebrity voices to create more powerful initiatives.

Explicit ethical standards and laws must govern the application of AI voice cloning. These standards should prioritize authorization, confidentiality, and disclosure to guarantee fair utilization. To create these regulatory structures and ensure that technological innovation is used to improve society while protecting against potential misuse or damage, industry executives, legislators, and ethicists must collaborate. This ethical application of AI voice cloning may open the door to a time when technological advances improve many facets of human existence and increase human capabilities.

## **6. Conclusion**

Artificial Intelligence Voice Cloning Generator is a game-changing technique that completely changes convenience, entertainment, and engagement. Its remarkable accuracy in recreating voices has paved the way for innovation in various industries, enabling tailored experiences and simplifying content production. It helps people with speech problems in the medical field by providing a more organic communication method. Additionally, it provides an affordable means of creating content and attracting viewers in the entertainment and advertising industries. Unquestionably, the developments in AI voice cloning represent a breakthrough in human-computer interaction and open the door to a society with more connectivity.

However, this innovation also raises relevant questions regarding how technology might be abused and the broader moral ramifications for society. Especially in the world of social media these days. The ease with which AI can mimic voices casts doubt on the veracity of audio recordings, which could make it more difficult to distinguish fake content from real stuff. Misuse of this technology to create deepfake audio for

malevolent purposes—like disseminating false information or impersonating someone—seriously threatens people's security, privacy, and the veracity of shared knowledge. Robust ethical frameworks, laws, and public awareness are essential to reducing these hazards and preventing the exploitation of AI voice cloning as it develops.

AI voice cloning emerges as a potent game-changer with the potential to transform connectivity, entertainment, and interpersonal relationships completely. However, because technology may be used as a tool for fraud and manipulation, it needs a well-rounded strategy that highlights its advantages while putting strict controls in place to deal with the moral and security risks currently facing humanity. We can only realize its promise and lessen the hazards it poses to society by ethical development practices, raising public awareness, and enforcing legislation.

## 7. Funding

This research paper received no internal or external funding.

## 8. Acknowledgments

The author thanks the Journal of Emerging Technologies (JET) and Jozac Publishers for the generous offer and opportunity to publish his research paper in the journal. Thank you so much!

## ORCID

Genesis Gregorious Genelza  <https://orcid.org/0000-0001-5577-7480>

## References

1. Aguilod, C. A., Tanduyan, M., Trangia, E., & Genelza, G. (2023). The learning interest in social studies education: In the lens of junior high school students. *Galaxy International Interdisciplinary Research Journal*, 11(9), 108-122.
2. Arik, S., Chen, J., Peng, K., Ping, W., & Zhou, Y. (2018). Neural voice cloning with a few samples. *Advances in neural information processing systems*, 31.
3. Babu, G. (2023). Embracing AI voice cloning technology. LinkedIn.com. Retrieved from <https://www.linkedin.com/pulse/embracing-ai-voice-cloning-technology-ganga-babu-ugeic>
4. Batalla, N. C., Onan, J. C. D., Tano, R. A., & Genelza, G. G. (2023). Virtual delivery of elementary teachers in the new normal: practices and implementation. *Galaxy International Interdisciplinary Research Journal*, 11(12), 229-248.
5. Bird, J. J., & Lotfi, A. (2023). Real-time Detection of AI-Generated Speech for DeepFake Voice Conversion. *arXiv preprint arXiv:2308.12734*.
6. Dale, R. (2022). The voice synthesis business: 2022 update. *Natural language engineering*, 28(3), 401-408.
7. Dewey, A., & Drahota, A. (2016). Introduction to systematic reviews: online learning module Cochrane Training.
8. El Zarzour, S. (2023). Caution Ahead: Exploring the Benefits and Dangers of AI Voice Cloning. *Telecom Review*. Retrieved from <https://www.telecomreview.com/articles/reports-and-coverage/7263-caution-ahead-exploring-the-benefits-and-dangers-of-ai-voice-cloning>
9. Espinosa, M. N. (2023). State of the art in Voice Cloning: A Review. Retrieved from <https://blog.marvik.ai/2023/03/21/state-of-the-art-in-voice-cloning-a-review/>
10. Genelza, G. G. (2022). Affective Language Learning and English Language Competence of Purposive Communication Students. *Universe International Journal of Interdisciplinary Research*, 2(10), 37-47.
11. Genelza, G. G. (2022). Internship program and skills development of fourth year bachelor of secondary education major in English. *Galaxy International Interdisciplinary Research Journal*, 10(2), 496-507.
12. Genelza, G. G. (2023). Quipper utilization and its effectiveness as a learning management system and academic performance among BSED English students in the new normal. *Journal of Emerging Technologies*, 3(2), 75-82.

13. González-Docasal, A., & Álvarez, A. (2023). Enhancing Voice Cloning Quality through Data Selection and Alignment-Based Metrics. *Applied Sciences*, 13(14), 8049.
14. Grand View Research (2023). AI Voice Cloning Market Size, Share & Trends Analysis Report By Component (Software, Service), By Deployment (On-premises, Cloud), By Application (Gaming, Advertising), By Vertical, By Region, And Segment Forecasts, 2023 – 2030 *Grandviewresearch.com*. Retrieved from <https://www.grandviewresearch.com/industry-analysis/ai-voice-cloning-market-report>
15. Polaris Market Research & Consulting. (2023). AI Voice Cloning Market: A Prospective Technique for Assimilating Distinct Vocal Features. *LinkedIn.com*. Retrieved from <https://www.linkedin.com/pulse/ai-voice-cloning-market-prospective>

