Kabarak Journal of Research & Innovation Vol. 2025 No. Supplement 1 (2025)
ISSN 2305-784X (print) ISSN 2410-8383 (online)
https://journals.kabarak.ac.ke/index.php/kjri/

Special Issue Sponsored by

Bridging Music and Media: AI-Driven and Human-based Solutions for Detecting Audio Deepfakes in an Era of Misinformation

Authors
Kadupe Sofola
Kabarak University
Email: kaymarxmusic@gmail.com

Benjamin Mbatia
Kabarak University
Email: bmbatia@kabarak.ac.ke

Amon Kirui
Kabarak University
Email: kipnoma@gmail.com

Article History: *Submitted: 29th October 2025; Accepted: 19th March 2026; Published (online): 11th August 2026*

Abstract

Artificial intelligence (AI) has become a vital part of the 21st-century digital world, serving as vital tools in enhancing human creativity and daily tasks. While it contributes to improving human lives, the rise of AI-generated audio deepfakes poses unprecedented challenges to both music and mass communication. One of the major challenges posed by AI in this regard is its contribution to the proliferation of misinformation through synthetic voices, manipulated interviews, and unreal musical performances. This paper explored AI-driven solutions for detecting deepfakes and examined human-based strategies in the same light, given the constant advancement of generative AI models and their ability to evade detection. The study employed a mixed-method approach. First, an experimental approach was used in creating deepfake music, which was later subjected to scrutiny by several AI-powered detectors. Secondly, a qualitative approach was used to source experts' recommendations on human-based strategies for combating deepfakes through available online interviews. Data was analyzed using content analysis. Findings revealed that existing AI-driven detection tools often misclassify or fail to identify AI-generated audio due to the continuous advancement in the development and updating of generative AI models. Human-based strategies such as critical listening and attention can help to identify robotic timbres, unnatural phrasing, the lack of breath, slips between frames that make the face jiggle, and linguistic imbalance in tonal

languages. Verification through official channels and record labels is also a key human-centered approach in detecting deepfakes and should complement the use of AI-based detectors. Future research should prioritize public perception of AI audio and the efficacy of media literacy across demographics, the cultural impact on music authenticity, the development of global regulatory frameworks, and the differential interpretation of multimodal versus audio-only deepfakes.

Keywords: AI deepfakes, audio forensics, misinformation, Music authenticity, detection tools

Introduction

AI tools are fast becoming an essential part of human lives, as new tools keep emerging, enhancing various aspects of human activities. Despite its advantages, lots of disadvantages have emerged in its use. For instance, a significant event occurred in April 2023, when an AI-generated song mimicking Drake and The Weeknd, 'Heart on My Sleeve,' went viral. The song had amassed millions of streams before being detected and pulled from platforms where it was uploaded. It pulled huge numbers on platforms such as TikTok, which had millions of views; Spotify (600,000 streams); Apple Music; and YouTube, which had pulled 275,000 views, before being removed due to copyright concerns that had been raised by Universal Music Group (UMG). The song had been created by an anonymous artist known as Ghostwriter. The event exposes the vulnerability of the music industry to audio deepfakes. Although the song was pulled down on streaming platforms, it had been downloaded by millions of listeners and had been in circulation across the internet as unofficial third parties continued to re-upload it (Torres, 2023; Grasser & Ruiz-Lichter, 2023).

Similarly, fabricated audio clips of politicians like the United States (U.S.) President Joe Biden have been weaponized to spread misinformation. During the build-up to the presidential election of the U.S. in early 2024, robocalls mimicking the voice of President Biden falsely admonished voters in New Hampshire not to participate in the primary election. This act was later traced to a political consultant who had perpetuated the act using AI voice-cloning technology (Yan et al., 2025). This illustrates how synthetic media threatens both cultural trust and democratic processes. Such events manipulate public perception, portraying public figures as saying what they had not said. Once such clips go viral on social media platforms, they spread so fast before fact checkers can intervene. The possibility of training such AI language models to create such content within a few seconds of being fed real audio calls for concern in the use of deepfakes for spreading misinformation. This concern comes from the possibility of bad actors possessing potent tools for sowing confusion and division through the mass media. Apart from undermining the long-existing trust in music and the media, these events raise ethical questions and legal concerns surrounding copyrights and the future of originality in music and the media.

Within the atmosphere of the growing synthetic media, there is a need to examine tools and strategies for detecting deepfakes in music and media. This study, therefore, intends to propose adaptive solutions for both music and media domains by exploring various strategies for detecting deepfakes in music versus spoken media. The study is important in various ways. First, it will provide potent insights to the multiple stakeholders grappling with the challenges of AI-generated audio deepfakes. For the music industry specifically, the study will address pressing concerns

about artistic authenticity and copyright protection within the atmosphere of expanding deepfake incidents like the viral "Heart on My Sleeve" deepfake track. Similarly, findings from the study will empower media professionals and fact-checkers in better detecting and combating synthetic audio misinformation, especially in such sensitive contexts as political elections, where fabricated clips can undermine democratic processes. In addition, this work bridges the disciplinary divide between music technology and mass communication, thereby offering comprehensive solutions to address both vocal manipulation in artistic works like music and speech forgery in news media. Lastly, the outcomes of the study will inform policy discussions, guide platform developers, and provide actionable strategies for content verification across industries.

Theoretical Framework: Media Ecology Theory

Media ecology theory, originally popularized by Marshall McLuhan in Understanding Media: The Extensions of Man (1964) and later advanced by Neil Postman in Amusing Ourselves to Death (1985), emphasizes that media and technology function as environments that shape human perception, social interaction, and cultural practices. McLuhan's well-known dictum, "the medium is the message," underlines the fact that the form of media itself, rather than the content it carries, has the most profound effect on society. Postman further argued that media technologies reconfigure cultural discourse, altering not only what we communicate but also how meaning is generated and consumed.

Applied to the present study, media ecology explains how AI-driven audio deepfakes disrupt the communicative environments of both music and mass media. In music, cloned voices and fabricated collaborations change how audiences relate to questions of authenticity, ownership, and originality. In journalism, synthetic voices that imitate political leaders contribute to disrupting the ecology of public trust and reducing the reliability of voice as an index of truth. This study's examination of detection tools and the emphasis on human-based measures in the detection of deepfakes may therefore be understood as efforts to rebalance a shifting media environment by embedding both technological correctives and human-based verification and detection tactics within it. As McLuhan (1964) noted, every medium creates both extensions and amputations; AI expands creative and communicative possibilities but simultaneously amputates authenticity. Media ecology thus provides a theoretical foundation for understanding the dual role of AI, both as a disruptor of traditional credibility and as a potential environment for new detection mechanisms that seek to safeguard authenticity in music and media.

Literature Review

This literature review explores three areas: AI in music, which focuses on deepfake vocals, AI in Mass Communication, focusing on synthetic speech in journalism/PR, and and a media ecology perspective to detecting deepfakes in audiovisual media, which delves into how the existence of AI-generated audio deepfakes affect the media's role.

AI in Music

The sudden influx of artificial intelligence into the creative spheres has produced both new and unpredictable opportunities and difficult issues, especially in music production. The recent scholarship demonstrates a growing contrast between technological innovation and artistic authenticity, as today in the emerging AI systems, innovative singing voices and musical compositions are being produced by computers, blurring the line between human and machine creativity (Li et al., 2024; Zhang et al., 2024). This giant leap forward in technology has compelled the music industry to face deep-seated questions regarding originality, copyright infringement, and the very nature of artistic expression. One of the central issues of the current research concerns singing voice deepfake detection (SVDD). As Guragain et al. (2024) note, contemporary AI vocal synthesis applications can mimic the voices of famous artists so well that initiatives such as the SVDD Challenge emerge to find schemes of reliable voice identification (Zhang et al., 2024). Approaches to SVDD often involve speech foundation model ensembles designed to enhance detection accuracy (Guragain et al., 2024).

The legal and ethical consequences of AI-generated music (AIGM) have become more contentious. Copyright paradigms developed over centuries have a hard time dealing with works that are not explicitly human-authored, hence a gray zone, as Fitas (2025) puts it, in the area of IP protection. This uncertainty has real consequences for artists, as can be seen by the viral nature of the AI-generated tracks that imitate the big artists while claims of copyright are still being enforced. Williams and Davidow (2025) persuasively warn that this technological change poses a threat of commodification of creative work at the expense of the human essence that brings music its cultural worth. Their concerns resound in the literature, such as Dall et al. (2024), who suggest ethical guidelines that would strive to find a balance between the innovating aspect and the artist protections. Assessment of the creativity of AI's output is another field of complexity. As Zhang et al. (2025) highlight, there is a constant gap between the creation of algorithmic music and actual human artistry, and the currently used evaluation measures usually miss the fine points that the listener's instinct identifies as "human." This gap implies that even though AI can imitate technical skills of composition, imitating the inner depths of emotion and the purpose behind human creativity is not the case.

The ramifications affect the wider realm of creative practitioners beyond music. Professional writers describe the advantages and worries concerning AI assistants, enumerating players' concerns about the removal of the individual voice and the creative agency (Ivanova et al., 2025). Comparative analysis of AI-generated versus human-generated content seeks to identify differences in writing styles, including complexity, semantics, readability, and structure (Sharma et al., 2025).

AI in Mass Communication

Similar to the way AI is revolutionizing the world of music, the mass media is also greatly impacted by the evolution of AI. Most notable in its enabling of unprecedented personalization, efficacy, content creation, and a new form of interaction between human and machine in enhancing output. All of these have led to deeper audience engagement. Despite the advantages offered by AI, studies have noted a series of challenges associated with its use, ranging from ethical to social

challenges. Studies such as Şenyapar (2024), Bormane and Blaus (2024), and Bhatlawande et al. (2024) have examined its application in the transformation of mass communication. For instance, Şenyapar (2024) noted in respect of marketing communication, highlighting the role of AI in enhancing personalized marketing and targeted advertisement, that AI is used to analyze large datasets, coming up with targeted messages that are employed in reaching the right audiences. Bormane and Blaus (2024) also buttress Şenyapar's (2024) position, emphasizing its impact on efficacy as a result of being able to analyze "high data volume" and facilitating instant, automated responses to queries.

Regarding communication in the public sector, Androutsopoulou et al. (2019) examined how AI-driven chatbots are helping to improve communications between governments and citizens, as well as between businesses and their customers. Scholars have also noted the impact of AI in facilitating speedy communication through smart replies and language-based models that translate between various languages and ensure real-time communication between individuals of different linguistic landscapes. Despite these, there is the concern that the use of AI can lead to emotional detachment between individuals and erosion of trust from negative perception of the use of AI, thereby posing a great challenge to authenticity (Hohenstein et al., 2021; Mieczkowski et al., 2021; Hohenstein & Jung, 2020; Purcell et al., 2024). In addition, despite the widespread use of AI in enhancing deliverables, Purcell et al. (2024) found that individuals often judge the use of AI by others in communication more harshly than their own. Lastly, scholars have also examined ethical considerations and literacy in the use of AI in mass communication. This stems from the mass personalization of communication content (Hermann, 2021; Saeed & Abdalla, 2025) and concerns regarding algorithmic bias, manipulation, and data security. This is, therefore, a growing call for AI literacy so that users can better understand how it works, thereby promoting responsible use (Mirek-Rogowska et al., 2024; Bormane & Blaus, 2024).

A Media Ecology Perspective to Detecting Deepfakes in Audiovisual Media

The disruptive impact of AI-generated audio deepfakes should not only be seen as a technological effect. Instead, the way the media culture and communication processes are influenced should also be acknowledged. For instance, the presence of deepfakes creates challenges to journalists' and editors' traditional gatekeeping roles (Voinea, 2025). Initially, journalists and editors controlled the flow of credible information (Simon, 2025). However, within the current digital environment, gatekeeping has become more complex. The reason is that synthetic voices usually bypass the existing editorial checks and circulate directly to audiences, allowing malicious actors to set agendas with fabricated events, even though it democratizes content creation. This happens through platforms, such as TikTok, YouTube, and X, whose information dissemination is largely decentralized.

In addition to the disruptive impact of AI-generated deepfakes on gatekeeping, the expectations of agenda-setting in media are also distorted. Particularly, the issues that are given prominence and how such actions influence what the public perceives as important have significantly changed (Kharvi, 2024). Instead of focusing on the main issues for discourse, the public's attention shifts to authenticity and manipulation (Boediman, 2025). Such extraneous discussions are caused by

the disproportionate coverage of viral AI songs or fabricated political speeches by the media. Therefore, instead of the public focusing on the real issues affecting society, as was the case traditionally, the debates are misdirected to technology regulation, an aspect that can require collaborative efforts with stakeholders from various sectors. Also, the legitimate media is forced to play the reactive role of debunking deepfakes, rather than proactively informing the public.

The existence of deepfakes also impacts framing, a critical media concept. The reason is that AI-generated deepfakes convey false information. Consequently, events, personalities, and narratives are strategically framed in ways that can shift public perception (Weikmann et al., 2024). For example, when an AI-generated song imitating a popular musician is created and shared online, the issue is not only framed as entertainment. Instead, existential topics of authorship and originality emerge in the public space when it is determined that indeed the audio is fake. Separately, political deepfakes do not advance important policy and governance issues. Instead, the framing diverts real issues into discourses on trust in leadership and institutions and potentially erodes democratic legitimacy (Neyazi et al., 2025). Consequently, the two-step flow of communication is disrupted. Instead of the media and influencers interpreting information first before it reaches the wider audience, deepfakes that are usually raw and manipulated reach the public directly and instantly through social media platforms. This way, the contextualization of media, which used to occur before consumption, is bypassed. Therefore, the media, being the source of all this information, becomes a subject of scrutiny in terms of how framing is managed.

The current popularity of AI-generated deepfakes also has an influence on how the importance of media literacy as a cultural defense mechanism is perceived. The ability to determine the authenticity of information has become an essential skill among media consumers (Diel et al., 2024). The reason is that every day, audiences are increasingly navigating spaces where the line between authentic and artificial blurs. Therefore, most media scholars are advocating that people should be able to evaluate sources critically to identify manipulation cues and verify information. The technique requires dedicated cross-checking of information before labeling it as genuine. Therefore, although media literacy was traditionally seen as a classroom concern, it has now expanded to a societal necessity that could be used to combat misinformation (López-Borrull & Lopezosa, 2025). This way, the credibility in music and mass communication can be preserved.

Finally, the theme of authenticity is another concern of media that should be reviewed in line with the generation and distribution of AI-generated audio deepfakes. The main issue here is the difference between representation and truth. Particularly, the human voice is regarded as a marker of identity and originality in music. The acknowledgement of the authentic human voice in music has been around for centuries. However, deepfake technologies have destabilized authenticity (Flattery & Miller, 2024). Consequently, artists and consumers face uncertainty in how to create and listen to music. The challenge can be summarized as one of the numerous redefinitions of cultural practices and communicative norms caused by the evolution of media, as posited by the media ecology perspective.

In sum, the existence of deepfakes threatens the generation of media through music or speech as some individuals become hesitant due to the fear of being manipulated or unable to distinguish

truth from fiction. Moreover, genuine media may be erroneously dismissed as deepfake, creating a crisis of signification in media. The implication is that the symbolic meaning of audio and video evidence becomes disregarded and viewed as no longer stable or trustworthy. Looking ahead in the literature, there is an imperative demand for interdisciplinary studies that directly target the technical, legal, social, and philosophical dimensions of the issue surrounding the use of AI in the mass media. It is for this reason that this study engages in an exploration of social and technical strategies in detecting deepfakes in music and media.

Methodology

The study employed a mixed-methods approach in evaluating AI-driven audio deepfake detection tools across music and spoken media. It employed an experimental approach in creating AI-generated music, which was subjected to scanning by AI tools for detecting deepfakes. It also employed a qualitative procedure by gathering additional data from expert interviews available on YouTube on strategies for detecting deepfakes. The interviews were selected using the purposive sampling technique, focusing on engagement between a media person and two experts in the industry: one, a computer scientist who specializes in creating fake videos of popular politicians, and the other, a professor and digital forensic expert. Apart from these, secondary data were collected from published academic materials relevant to the study. The analysis of data was carried out using content analysis. Ethical considerations include a declaration of the source of primary data

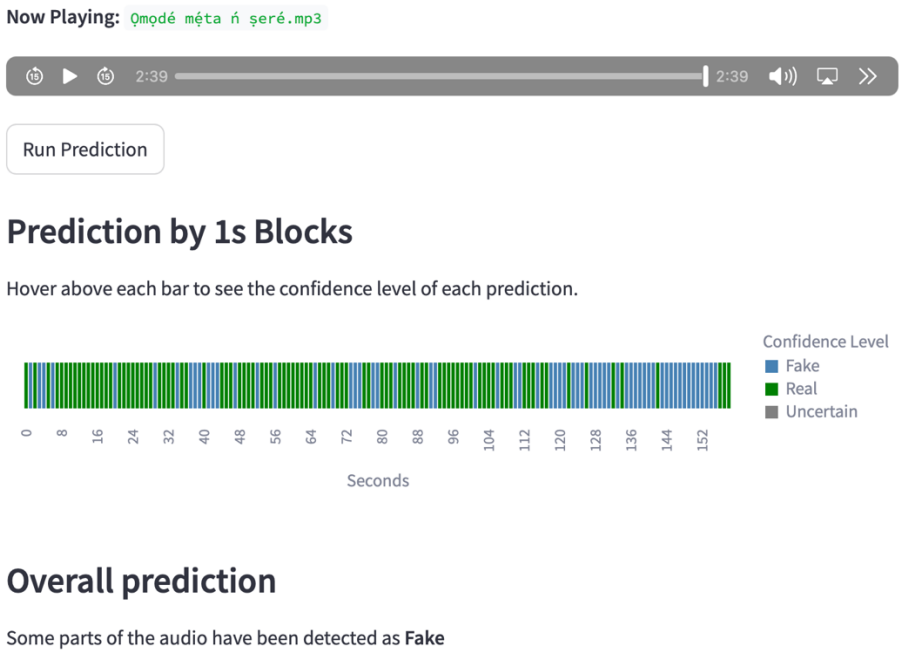
Findings

This section presents the findings obtained from the examination of the detection of deepfakes in musical vocals and audiovisual media. It details the results of the experiments conducted with various AI-generated audio and video content. The performance of several automated detection tools is evaluated, and their limitations are documented. The text also outlines human-based strategies for identifying synthetic media.

Detecting Deepfake Vocals

The evolution of generative AI tools has seen the rise of tools that can produce full musical tracks with a prompt, carrying out the entire process that would normally require humans to write a song, record the song in a studio to avoid unwanted sounds, add instrumental accompaniments, and mix and master the recording. For these AI tools, the long process that takes hours to accomplish by humans is achieved within seconds, provided that they are given the appropriate prompts. In an experimental endeavor, the authors generated a song track using Suno, a generative AI tool used in producing music, using human-written lyrics in the Yorùbá language. The Yoruba language was deliberately chosen due to its tonal nature that requires the logogenic principle in music composition and singing. In doing this, the prompt instructed Suno to create the music in the Nigerian Afrobeats style. The resulting content was a Nigerian Afrobeats track that might be difficult to detect as a deepfake, if not for the inability of the algorithm to align the Yorùbá lyrics with the rise and fall of the speech. In addition, the generated song was uploaded to Hugging Face,

a fake audio detection tool powered by Mozilla-AI. Although the tool was able to detect some fake elements in the generated track, it did most of the characteristic features of the song as "real" or human-generated. The result is shown below.



A further examination of the same audio using an audio deepfake detector powered by Justbuildthings provided more insightful analysis on the track. It identifies unnatural speech patterns, inconsistent voice characteristics, unusual transitions, and distortions. Inasmuch as this tool was able to provide detailed analysis of the audio file, it is unable to maintain a stance about whether the audio is real or fake. We took it further by analyzing a recorded live performance of Fela's "Lady," and the results were similar to the initial AI-generated song. This exemplifies the limitedness of existing detection tools in detecting AI-generated songs, especially those laced with propaganda and misinformation.

Given these identified limitations, it has become a reality that generative AI tools in this realm keep improving to the extent of generating almost-perfect songs and make detection become more difficult to the extent that detection tools become confused and even misclassify human-generated music as possibly generated by AI. In light of this, there are a few strategies that can help in detecting deepfake songs. The first is critical listening. Careful listening and attention to details can reveal robotic timbre, unrealistic phrasing, lack of noticeable breath, and linguistic imbalance, especially in tonal languages. Secondly, inasmuch as advancement in the sophistication of generative AI tools will continue to make it difficult for detection tools and even humans to detect deepfakes, verification will go a long way in doing so. For instance, streaming from verifiable platforms can, to a large extent, guarantee the prevention of deepfakes. Likewise, songs can be verified directly from record labels and their official channels, such as on YouTube. The music's architectural structure puts an identity to every sound and accompanying video. A deviation from

this structure should call for suspicion, prompting listeners to carry out a simple background check through the official streaming platforms and channels of the artist or record label.

When it comes to voice cloning and audio deepfakes, it has a lot to do with music and the media. An instance is given earlier about an AI-generated song that mimicked Drake and The Weekend's "*Heart on My Sleeve,"* which had amassed millions of streams before it was pulled down for copyright-related reasons. Additionally, AI was used to create a fake podcast with Joe Rogan and Steve Jobs, bringing back Steve Jobs, who had been dead for ten years, in what sounded convincing, should Jobs still be alive at the time the podcast was released. Several such events have also occurred ever since, raising concerns about the future of authenticity regarding mass media. While clues like the absence of noticeable breath patterns, slight phrasing glitches, unnatural rhythmic patterns, and comparison with verified audio recordings of an individual or artist can help to a large extent, reliance on verified, trusted channels seems to be the most viable strategy in combating the spread of deepfake audios.

Detecting Deepfakes in Audiovisual Media

Although various tools continue to emerge for the purpose of detecting deepfakes in the media, it is important to note that AI-generated deepfakes are becoming more convincing and more difficult to detect. This is because developers of AI generative tools are not hanging up their boots. Rather, they are working harder to improve on their inventions, such that they are able to generate videos, sounds, images, and texts that are as convincing as possible and can be undetectable by both humans and detection tools. For instance, while tools have been built to effectively detect AI-written texts, the same tools, on many occasions, are able to "humanize" the text, such that it reads 0% AI content when scanned on an AI-detecting tool. This is an indication that the realm of artificial intelligence is an evolving world, where detection tools alone may not be sufficient in deciphering deepfake content, especially in the mass media. The mass media is unique, basically because of its heavy reliance on audio-visual content, which is more convincing than audio or print media.

While there are various detection tools like Deepware Scanner, Resemble Detect (Resemble AI), RawNet2 (Academic Model), and Wav2Vec 2.0–based Detection Models, all of which are not 100% accurate in detecting deepfakes. Their accuracy rate depends largely on the type of deepfake and the tool used in creating them, training data that limits their functionality, resolution and compression, and adversarial adaptation as developers improve and block loopholes. Hao Li, an associate professor of computer science who has created nearly perfect deepfake videos of politicians, noted in an interview with Janina Semenova of DW Shift that manipulations and deepfakes are getting better, and it is only a matter of time before deepfakes become almost undetectable by both humans and deepfake detection tools.

In a comprehensive analysis of deepfake detection models which focuses on accuracy, generalization and adversarial attacks, Abbasi et al. (2025) employed an experimental approach in evaluating three leading deepfake detection models – XCeption, ResNet, and VGG16. Their findings reveal that although these tools have proven to be reliable, none of them is absolutely

accurate. XCeption, which scored the highest in detection rate, achieved an accuracy rate of 89.2% on the Deepfake Detection Challenge, with strong generalization and real-time suitability, which makes it effective to a large extent in detecting deepfakes, especially in live video and social media monitoring. One major challenge identified by this experiment is adversarial attack, which entails the use of deliberately manipulated data designed to mislead deepfake detection models. More specifically, this is carried out in the form of subtle, strategically crafted perturbations that are usually difficult to be perceived by the human eyes or ears. This not only deceives the human eyes; it also misleads detection algorithms, making them misclassify deepfakes as authentic or, sometimes, authentic videos as deepfakes.

Despite this challenge, however, industry experts have clamored for human intelligence as a major requirement in the detection of deepfakes. This is specifically important in the African context and most of the Global South, given that the technologies behind generative AI are imported from the Global North. In addition, high-end deepfake detectors usually require a subscription, which might pose an accessibility challenge to many industry players. Hany Farid, a digital forensic expert and professor at the University of Berkeley, noted human-based strategies that can help in detecting deepfakes in an interview with Janina Semenova of DW Shift. According to Farid, such strategies include analyzing various videos of politicians in an attempt to study their facial characteristics when they talk. This includes how they move their mouth and face during speeches and how their expressions change. It also includes a comparison of the movement of the eyes based on existing verified videos. This method is meant to detect anomalies in fake videos and it is entirely human-based as it does not involve the use of any AI tool. It is based on the assumption that current AI models are vulnerable in those aspects at the moment, and since it is one of their major areas of weakness, it provides valuable clues for human detection in forensic analysis of deepfakes.

Further, Farid revealed that deepfakes are made frame after frame, with weak spots such as slips between the frames, making the face jiggle a little. However, to detect such weakness, the videos have to be played at a very slow speed in such a way that the frames become obvious and these glitches noticeable. Although Farid, from an expert perspective, has outlined these human-based strategies, he admits the possibility of deepfakes becoming better at addressing these weaknesses. This, therefore, becomes a major concern—while Farid speaks from the position of a forensic expert, his admission of the possibility of deepfakes outsmarting detectability makes valid the position of Hao Li, that in the long run, the most viable option in detecting deepfakes would be for people to trust the source where media content comes from. There is also the reliance on common sense in determining what is plausible, especially given the antecedents of individuals represented in such content and the codes guiding their office or the capacity in which they serve. In addition, there is a need for media houses to be as responsible as possible so that they do not tamper with their credibility, given that they remain a major factor in preventing the spread of deepfakes.

Conclusion

It has been demonstrated by this study that the rise of AI-generated audio deepfakes presents profound implications for both music and mass communication. From viral tracks such as Heart on My Sleeve to fabricated political messages, synthetic media is reshaping cultural trust,

democratic discourse, and the perception of authenticity. The findings reveal that although AI-driven detection tools provide partial solutions, they remain limited in accuracy due to adversarial adaptation and the rapid evolution of generative models. The misclassification of synthetic content as "real" and the struggle with linguistic nuance by AI-powered detection tools affirm their critical vulnerability in catching up with new-generation deepfake techniques and adversarial attacks. The research results are congruent with the common media studies perspective of technological determinism, which asserts that institutional safeguards and social norms are constantly and continuously challenged by technological change. The study has further indicated that human-centered detection strategies are equally important. Individuals can seek to validate the authenticity of audio messages by listening critically, cross-verifying with official channels, and paying forensic attention to audiovisual inconsistencies. This approach is consistent with the gatekeeping role of media professionals of ensuring that, in addition to their traditional responsibilities, they should filter and verify information to combat deepfake manipulations. Considering the findings on AI deepfakes, the study contributes to public awareness, perception of risk, and policy responses, and this corresponds with the agenda-setting function of mass media. The framing of authenticity in media and music is challenged by the propagation of synthetic audio content. The reason is that audiences used to rely on voice as a trustworthy index of presence and originality, but the mediation of AI now blurs the boundaries between real and artificial, and this echoes Postman's concerns about technology reshaping cultural discourse.

Recommendations

Considering the study's findings, media literacy is recommended as the ultimate deepfake counter method because both audiences and media organizations should learn how to critically consume and proactively verify content, respectively. A hybrid approach would be necessary in upholding communication integrity during this digital age. Particularly, AI-based detection mechanisms should be combined with human-based scrutiny, institutional responsibility, and informed audiences. As the media ecology framework suggests, every medium simultaneously extends human capacity and amputates traditional assumptions. AI extends creativity and efficiency but amputates trust and authenticity. Therefore, it will be essential to embed resilience into technological, cultural, and communicative practices. In addition to technical detection, democratic values, cultural heritage, and the credibility of both music and journalism should be religiously preserved in a world that has become increasingly mediated by artificial intelligence. However, the path forward requires a concerted effort from technologists, artists, journalists, policymakers, and educators. There is a need for regulations that require the disclosure of AI-generated audio content by compelling individuals and media organizations to embed watermarks or metadata tags into all synthetic files. AI impersonations of artists' voices could be curbed through copyright and intellectual property protection. Social media platforms should be forced to create robust policies against malicious deepfake circulation. In the future, studies should focus on the perception of AI-generated audio and the effectiveness of media literacy programs in improving detection by different demographic groups. Other areas of focus could be on the impact of the erosion of authentic voice in music in reshaping cultural values, fan relationships, and trust in media institutions; regulation of audio deepfakes and best practices adopted by different nations; and the differential interpretation of multimodal fakes and audio-only content.

Funding Statement

This research received no external funding.

Conflict of Interest Statement

The authors declare that there is no conflict of intertest.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

References

Abbasi, M., Váz, P., Silva, J., & Martins, P. (2025). Comprehensive Evaluation of Deepfake Detection Models: Accuracy, Generalization, and Resilience to Adversarial Attacks. *Applied Sciences*, *15*(3), 1-16. https://doi.org/10.3390/app15031225

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, *36*(2), 358-367. https://doi.org/10.1016/J.GIQ.2018.10.001

Bhatlawande, S., Patil, T., & Anavatti, A. (2024). AI in communication. *EPRA International Journal of Multidisciplinary Research (IJMR)*, *10*(11), 400-401. https://doi.org/10.36713/epra19105

Boediman, E. P. (2025). Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review. *Jurnal Komunikasi*, *19*(2), 313–334. https://doi.org/10.20885/komunikasi.vol19.iss2.art8

Bormane, S., & Blaus, E. (2024). Artificial intelligence in the context of digital marketing communication. *Frontiers in Communication*, *9*, 1-13. https://doi.org/10.3389/fcomm.2024.1411226

Dall, M., Herzog, K., Hufnagel, A., Ibsen, D. B., Lebiecka-Johansen, B., Ruppert, P. M. M., Preston, J. M., Toh, P. J. Y., & Yfanti, C. (2024). Our future, we decide: five ways to reform the scientific publication process. *Nature Reviews Endocrinology*, *21*(1), 5–6. https://doi.org/10.1038/s41574-024-01056-x

Diel, A., Lalgi, T., Schröter, I. C., MacDorman, K. F., Teufel, M., & Bäuerle, A. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers.

Computers in Human Behavior Reports, *16*(1), 1–13. https://doi.org/10.1016/j.chbr.2024.100538

DW Shift. (2019, December 20). *How to detect deepfakes | Deepfakes explained* [Video]. YouTube. https://www.youtube.com/watch?v=BuufkPTFt0E

Fitas, R. (2025). *The author is sovereign: A manifesto for ethical copyright in the age of AI.* http://dx.doi.org/10.48550/arXiv.2504.02239

Flattery, T., & Miller, C. B. (2024). "Deepfakes and dishonesty." *Philosophy & Technology*, *37*(4), 1–24. https://doi.org/10.1007/s13347-024-00812-1

Grasser, J. & Ruiz-Lichter, S. (2023). Ghostwriter in the machine: Copyright implications for AI-generated imitations, *National Law Review*. https://natlawreview.com/article/ghostwriter-machine-copyright-implications-ai-generated-imitations/

Guragain, A., Liu, T., Pan, Z., Sailor, H. B., & Wang, Q. (2024, December). Speech foundation model ensembles for the controlled singing voice deepfake detection (ctrsvdd) challenge 2024. In *2024 IEEE Spoken Language Technology Workshop (SLT)* (pp. 774-781). IEEE.

Hermann, E. (2021). Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective. *New Media & Society*, *24*, 1258 - 1277. https://doi.org/10.1177/14614448211022702

Hohenstein, J., & Jung, M. (2020). AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behaviour, 106*, 1-15. https://doi.org/10.1016/j.chb.2019.106190

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., ... & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, *13*(1), 5487. https://doi.org/10.1038/s41598-023-30938-9

Ivanova, A., Fedorova, N., Tilga, S., & Artemova, E. (2025). *Surveying professional writers on AI: Limitations, expectations, and fears*. 1–17. http://dx.doi.org/10.48550/arXiv.2504.05008

Kharvi, P. L. (2024). Understanding the impact of AI-generated deepfakes on public opinion, political discourse, and personal security in social media. *IEEE Security & Privacy*, *22*(4), 115–122. https://doi.org/10.1109/msec.2024.3405963

Kubovics, M. (2025). AI In the Creation of Marketing Communication Content. *Interantional journal of scientific research in engineering and management*, *9*(4), 1-8. https://doi.org/10.55041/ijsrem42613

Li, Y., Milling, M., Specia, L., & Schuller, B. W. (2024). From audio deepfake detection to AI-generated music detection -- A pathway and overview. *Cornell University*. https://doi.org/10.48550/arxiv.2412.00571

López-Borrull, A., & Lopezosa, C. (2025). Mapping the impact of generative AI on disinformation: Insights from a scoping review. *Publications*, *13*(3), 33–50. https://doi.org/10.3390/publications13030033

McLuhan, M. (1964). *Understanding media: The extensions of man*. New York: McGraw-Hill.

Mieczkowski, H., Hancock, J., Naaman, M., Jung, M., & Hohenstein, J. (2021). AI-Mediated Communication. *Proceedings of the ACM on Human-Computer Interaction*, *5*, 1 - 14. https://doi.org/10.1145/3449091

Mirek-Rogowska, A., Kucza, W., & Gajdka, K. (2024). AI in communication: Theoretical perspectives, ethical implications, and emerging competencies. *Communication Today*, *15*(2), 16-29. https://doi.org/10.34135/communicationtoday.2024.vol.15.no.2.2

Neyazi, T. A., Khai Ee, T., & Kuru, O. (2025). Campaign deepfakes and affective polarization: The role of artificial intelligence in campaigns in shaping voter attitudes. *Social Science Computer Review*, *1*(1), 1–16. https://doi.org/10.1177/08944393251362247

Postman, N. (1985). *Amusing ourselves to death: Public discourse in the age of show business*. Viking Penguin.

Purcell, Z. A., Dong, M., Nussberger, A., Köbis, N., & Jakesch, M. (2024). People have different expectations for their own versus others' use of AI-mediated communication tools. *British Journal of Psychology*, *1*(1), 1–19. https://doi.org/10.1111/bjop.12727

Saeed, S., & Abdalla, O. (2025). AI and its impact on communication. *Journal of English Language Teaching and Applied Linguistics*, *7*(2), 88-94. https://doi.org/10.32996/jeltal.2025.7.2.10

Senyapar, H. N. D. (2024). Artificial intelligence in marketing communication: A comprehensive exploration of the integration and impact of AI. *Technium Social Sciences Journal*, *55*(1), 64–81. https://doi.org/10.47577/tssj.v55i1.10651

Sharma, T., Sachdev, P., Priya, & Kumari, S. (2025). Unveiling writing styles: A comparative analysis of AI-generated and human generated content. *AIJR Proceedings*, 193–212. https://doi.org/10.21467/proceedings.178.22

Simon, F. M. (2025). Rationalisation of the news: How AI reshapes and retools the gatekeeping processes of news organisations in the United Kingdom, United States and Germany. *New Media & Society*, *1*(1), 1–21. https://doi.org/10.1177/14614448251336423

Torres, M. (2023). (A.I.) Drake, The weeknd, and the future of music. *Washington Journal of Law, Technology & Arts*. https://wjlta.com/2023/05/03/a-i-drake-the-weeknd-and-the-future-of-music/

Tumiran, M. A., Mohammad, N., & Bahri, S. (2024). Implications of leveraging AI in students' academic writing: A review analysis. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, *9*(8), e002954. https://doi.org/10.47405/mjssh.v9i8.2954

Voinea, D. V. (2025). Reconceptualizing gatekeeping in the age of artificial intelligence: A theoretical exploration of artificial intelligence-driven news curation and automated journalism. *Journalism and Media*, *6*(2), 68–86. https://doi.org/10.3390/journalmedia6020068

Weikmann, T., Greber, H., & Nikolaou, A. (2024). After deception: How falling for a deepfake affects the way we see, hear, and experience media. *The International Journal of Press/Politics*, *30*(1), 187–210. https://doi.org/10.1177/19401612241233539

Williams, P., & Davidow, S. (2025). Humans versus robots: Countering AI and the commodification of creative writing. *New Writing*, *22*(2), 1-15. http://dx.doi.org/10.1080/14790726.2025.2489356

Yan, H. Y., Morrow, G., Yang, K. C., & Wihbey, J. (2025). The origin of public concerns over AI supercharging misinformation in the 2024 US presidential election. *Harvard Kennedy School Misinformation Review*, *6*(1), 1-13. http://dx.doi.org/10.31219/osf.io/cpy7v

Zhang, H., Liang, J., Phan, H., Wang, W., & Benetos, E. (2025). From aesthetics to human preferences: Comparative perspectives of evaluating text-to-music systems. *arXiv preprint arXiv:2504.21815*. https://doi.org/10.48550/ARXIV.2504.21815

Zhang, Y., Zang, Y., Shi, J., Yamamoto, R., Toda, T., & Duan, Z. (2024, December). SVDD 2024: The inaugural singing voice deepfake detection challenge. In *2024 IEEE Spoken Language Technology Workshop (SLT)* (pp. 782-787). IEEE. http://dx.doi.org/10.48550/arXiv.2408.16132