DOI https://doi.org/10.30525/2592-8813-2024-1-11

OPEN SOURCE AND PROPRIETARY SOFTWARE FOR AUDIO DEEPFAKES AND VOICE CLONING: GROWTH AREAS, PAIN POINTS, FUTURE INFLUENCE

Vitaliy Danylov,
Graduate Student, MET Department of Computer Science,
Boston University (Boston, USA)
ORCID ID: 0009-0004-8919-3378
vitaliy_danylov@ukr.net

Abstract. The article is dedicated to exploring the rapidly growing field of audio deepfake and voice cloning technologies. It examines the dual pathways of development in open-source and proprietary software, identifying key areas of growth, challenges faced by developers and users, and the potential impact these technologies may have on various sectors. The relevance of this article lies in its timely examination of a rapidly evolving technology that has significant implications for privacy, security, and content authenticity in the digital age. The research results show that audio deepfakes signify a major leap forward in our ability to generate and modify audio recordings, enabling the creation of highly convincing imitations of specific voices. This technology spans three main categories: imitation-based, synthetic-based, and voice cloning, each offering unique applications and introducing distinct challenges. These advancements have opened up new possibilities in fields such as entertainment, customer service, and security but also bring to light serious ethical and security considerations. The imperative for careful oversight and the development of regulatory frameworks to prevent misuse is clear. The ecosystem of audio deepfake technology features a wide range of open-source and proprietary software, each designed to meet specific requirements. Prominent solutions like Resemble.ai, Descript, and CereProc, among others, cater to diverse needs from entertainment to multilingual voice cloning. Alongside these, open-source projects like FaceSwap and Real-Time Voice Cloning offer valuable resources for innovation, emphasizing the importance of responsible usage and ethical development. The trajectory of audio deepfakes is marked by both promising prospects and formidable challenges. The potential for these technologies to revolutionize storytelling, create personalized experiences, and support educational initiatives is immense, facilitated by ongoing advancements in AI and the growth of open-source communities. However, the concerns surrounding ethical use, the demand for computational resources, and the challenge of achieving linguistic diversity underline the need for comprehensive ethical guidelines and sophisticated detection mechanisms. Navigating the future of audio deepfakes will require a balanced approach, weighing their transformative potential against the risks they pose. The practical significance of audio deepfakes extends into research and development, where they can be used to study speech disorders, aid in voice restoration for individuals who have lost their ability to speak and explore new forms of human-computer interaction. As the technology matures and becomes more accessible, its practical applications are expected to expand, potentially transforming how we interact with digital content and each other in virtual environments.

Key words: open source, audio deepfakes, voice cloning, software.

Introduction. The trajectory of deepfake technology development is closely intertwined with broader trends of the Fourth Industrial Revolution, characterized by the convergence of digital, biological, and physical worlds and the emergence of breakthrough technologies such as artificial intelligence.

The current digital information landscape witnessed a rapid expansion of the artificial intelligence sector: in 2021, the AI applications market was valued at 93.5 billion USD, and today, it is estimated at 245 billion USD, and by 2030, it is projected to reach 1.7 trillion USD, accounting for an annual growth rate of 38%. AI-based communication technologies, particularly audio deepfakes and voice cloning, are gaining particular relevance (Whittaker et al., 2023).

Audio deepfake technology, as a notable application in the AI-driven communication sector, can blur the lines between reality and fiction, creating both innovative opportunities and severe chal-

lenges in various fields, especially on online communication platforms where deepfakes are becoming increasingly prevalent.

Industry reports indicate a sharp increase in fake content: over 100 million fake videos circulated online in 2020, which, between 2017 and 2020, amassed over 5.4 billion views on YouTube alone. This growing presence underscores the increasing influence of deepfakes on the digital communication landscape (Whittaker et al., 2023).

Against this backdrop of rapid growth, an academic and practical discourse around audio deepfakes is developing, with new discussions emerging about their potential applications and the ethical considerations they bring. However, a consensus on these technologies' future, challenges, and promising applications is not yet clearly defined. To understand the future development of the audio content sphere, it is essential to examine expert and scientific research worldwide, as innovations can come from any direction. Z. Almutairi and H. Elgibreen (2022) highlight modern methods for detecting audio deepfakes, focusing on the challenges and future directions in this field and indicating the need to develop effective tools to combat audio manipulations. Z. Khanjani, G. Watson, and V. Janeja (2023) also explore audio deepfakes, reviewing existing approaches and challenges in this area. M. Broz (2023) and T. Kamunya (2023) focus on analyzing software for creating audio deepfakes, including open-source tools, and highlight the top tools available in the market. It allows for assessing the potential and limitations of existing technologies in this field. Authors such as B. Fauve (2023) and E. Hays (2023) consider the broader implications of audio deepfakes, including societal and ethical aspects of these technologies, emphasizing the importance of understanding the potential impact of audio manipulations on various aspects of life. Researchers such as A. Naitali, M. Ridouani, F. Salahdine, and N. Kaabouch (2023), as well as L. Whittaker, R. Mulcahy, and K. Letheren (2023), investigate methods for generating and detecting deepfakes, highlighting the need for developing effective means of protection against manipulative audio materials. J. Kietzmann, A. Mills, and K. Plangger (2020) examine the future of advertising and branding in the context of deepfakes, pointing to potential changes in the perception of reality and its impact on consumers.

As deepfake technologies continue to evolve, they represent a significant potential for innovative research in the current era.

Main part. The purpose of the article is to provide a comprehensive overview of the software and technologies used in the field of audio deepfakes and voice cloning, and to highlight the main challenges and future prospects of these technologies.

To achieve this goal, the following points will be covered:

1) define what audio deepfakes are and their types;
2) review the main software used for their creation;
3) determine the significance of open-source technologies and list the programs that have been developed to date;
4) highlight the issues associated with the use of audio deepfakes;
5) identify the prospects of this technology.

Results and discussions. In recent years, the digital content world has seen the development of technologies capable of reproducing authentic voices, including imitating, cloning, and distorting them. Initially created for benevolent purposes, these technologies have also been illicitly appropriated for spreading false information globally via audio media. Such misuse has raised concerns about "audio forgeries," also known as sound alterations, which can now be easily created using ordinary smartphones or desktop computers. This development has sparked global concern regarding cybersecurity, highlighting the potential negative consequences of using audio forgeries. Despite this innovation's advantages, Audio Deepfakes (AD) extend far beyond simple textual communication or hyperlink exchange. They can be utilized for voice-based logical access forgery schemes, potentially affecting public sentiments through propaganda, defamation, or terrorist acts. Given the vast volumes

of voice data transmitted daily over the Internet, distinguishing authentic recordings from fraudulent ones is challenging (Almutairi & Elgibreen, 2022).

The definition of audio deepfakes. Audio deepfakes are artificially created or manipulated audio recordings that convincingly mimic a specific individual's voice, creating the illusion that they are saying something they actually did not. Initially developed for business purposes, such as enhancing the realism of audiobooks by reproducing soothing human voices, this technology has found more controversial applications.

There are three main categories of audio forgeries, each distinguished by the method of creation:

1) Imitation-based forgeries involve transforming one voice into another to create the impression that another person is speaking. It can be achieved in various ways, including engaging individuals who can accurately imitate the target speaker's voice. More sophisticated methods involve using digital algorithms, such as Efficient Wavelet Masking (EWM), to alter the characteristics of the original audio signal to match the target voice. This process involves recording both the original and the target voice with similar characteristics and then transforming the original audio to replicate the speech patterns of the target voice, resulting in a forged audio clip that, to an untrained ear, might sound almost indistinguishable from a genuine recording (Almutairi & Elgibreen, 2022).

2) Synthetic-based forgeries: Text-to-speech (TTS) forgeries are generated by converting text into realistic spoken audio. The process typically involves three main stages: text analysis, where the input text is converted into linguistic features; acoustic modeling, which uses these features to generate speech sounds; and finally, a vocoder, which synthesizes actual speech. Advanced TTS systems like Tacotron 2, Deep Voice 3, and FastSpeech 2 are known for creating natural-sounding speech. These systems require a database of clean, structured audio recordings and corresponding textual transcripts to train the model, which generates the synthetic voice (Almutairi & Elgibreen, 2022).

3) Voice cloning: This category involves the direct reproduction of the recorded speech of the target individual, often used in malicious contexts to deceive listeners. There are two main subtypes within this category: far-field detection, where the target individual's voice recording is played through a device, such as a telephone handset, to simulate a live conversation, and cut-and-paste detection, where different audio segments of the victim's speech are stitched together to create a new, fraudulent message (Almutairi & Elgibreen, 2022).

Depending on their specific objectives, various types of audio deepfakes are employed. For instance, imitation-based deepfakes might be used in entertainment to protect the privacy of an original voice by transforming it to sound like another. Synthetic-based or Text-To-Speech deepfakes could be utilized in customer service to provide real-time, natural-sounding responses. Replay-based deepfakes might be used in security testing to assess the robustness of voice recognition systems.

Software for creating audio deepfakes. Software for creating audio forgeries employs advanced technologies to generate synthetic sound that mimics real voices with remarkable accuracy. Despite its fraudulent uses, this innovation has immense potential in various fields, transforming content creation, marketing, and branding. Let's consider the main advantages and prospects of using modern technologies for creating audio forgeries.

In the rapidly changing landscape of the business world, a multitude of software solutions are continuously being developed and deployed to meet evolving business needs. As of 2024, the realm of audio manipulation and creation has seen significant advances, with a variety of tools gaining prominence for their innovative features. Among these, notable software includes Resemble.ai, renowned for its applications in the entertainment industry; Descript for its podcasting and audio editing capabilities; and CereProc, distinguished for its multilingual voice cloning. Each tool is tailored to specific industry requirements, offering solutions from fast voice synthesis with real-time voice cloning to specialized applications in video games and voice-over projects with Replica Studios (Broz, 2023).

Table 1

The main advantages and perspectives of using modern technologies for creating audio deepfakes (Kamunya, 2023; Kietzmann et al., 2020)

Advantages	Overview
Innovative Storytelling	Enables more creative storytelling techniques in podcasts and audiobooks.
Efficient Content Updates	Allows easy updates to existing audio content, saving time and resources.
Educational Tools	Enhances language learning and educational materials with realistic voice simulations.
Research and Development	Assists in linguistic and psychological studies by providing diverse audio samples for analysis.
Personalized Customer Experiences	Voice cloning can enable brands to offer highly personalized experiences to their customers. For instance, using a customer's preferred voice for virtual assistants or customer service bots can make interactions more engaging and comfortable, leading to increased customer satisfaction and loyalty.
Brand Mascots and Voices	Brands can create or clone distinctive voices for their mascots or spokespersons, ensuring consistency across various marketing channels. This not only strengthens brand identity but also improves brand recognition among consumers.
Multilingual Content Creation	Audio deepfake technologies facilitate the creation of multilingual content without the need to employ multiple voice actors and native language speakers. This can significantly expand a brand's global reach and resonance with diverse audiences while maintaining a consistent brand voice.

Let's look at the top 10 programs and online applications that allow you to create audio products based on audio deepfakes.

Table 2

Top applications that allow the creation of audio products based on audio deepfakes (Broz, 2023)

Program	Focus Area	Price Range
Resemble.ai	Entertainment	$99/month – Custom
Descript	Podcasting & Audio Editing	Free – Custom
CereProc	Multilingual Voice Cloning	$499.99
Respeecher	Filmmaking & Video Game Design	$200/month
iSpeech	Customer Service & Video Game Design	Not Specified
ReadSpeaker	Business Narration	Custom
Fliki AI	Voiceovers and Voice Cloning	$21-66/month
ElevenLabs's Voice Changer	Fun & Entertainment (Mobile)	Free-$22/month
Speechify	Audiobooks	Free – $139/month
Wavel AI	Podcasting & Voiceovers	$18-60/month

Note: systematized by the author

Given that the technologies listed in the table are proprietary and require payment, audio content creation professionals might also want to test free technologies. Open-source technologies serve this need by offering tools anyone can use, modify, and distribute. These technologies are built on the principle of community collaboration, enabling continuous improvement and innovation while maintaining transparency and ethical standards in their development and application. Thus, open-source

platforms become invaluable resources for professionals who seek cost-effective, adaptable solutions for audio deepfake creation and manipulation.

The open-source technologies mentioned below focus on various audio and visual deepfake creation aspects. For example, Real-Time-Voice-Cloning specializes in cloning voices to produce speech from text, which is suitable for creating synthetic audio content. These tools offer innovative solutions for generating realistic audio and visual deepfakes, each with unique features catering to different needs within the deepfake creation spectrum (Kamunya, 2023).

Let's take a look at the list of open-source applications that are used to create audio deepfakes.

Table 3

Top open-source applications used to create audio deepfakes (Kamunya, 2023)

Program	Description	Application Area
1	2	3
CorentinJ's Real-Time-Voice-Cloning	Real-Time-Voice-Cloning is written in Python and implements Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis (SV2TTS) deep learning framework for Windows and Linux. It creates a digital representation of a voice from a few seconds of audio in English, and then this representation is used as a reference to generate speech from text.	Able to cover voiceovers and synthetic audio content generation in projects where multilanguage support is not required
RVC-Boss's GPT-SoVITS	GPT-SoVITS is a voice conversion and text-to-speech framework written in Python that currently covers Japanese, Chinese, and English. It allows the launch of voice cloning model training with just one minute of audio data and supports further model training and fine-tuning, to improve voice similarity and realism.	Currently targets video game voiceovers and the development of virtual assistants providing customer support in English and Chinese
Wunjo AI	Wunjo AI is an advanced speech & deepfake neural network tool written in Python and JavaScript. It synthesizes and clones voices in English, Russian, and Chinese. This framework provides real-time speech recognition, as well as additional visual features like deepfake face and lips animation and deepfake face swaps.	Likely employed in synthetic audio creation for diverse multimedia uses
eSpeak NG	eSpeak NG is a compact speech synthesis engine written in C and Java for Windows, Linux, MacOs, and Android, supporting more than 100 languages and accents, therefore, catering to a diverse user base. The synthesized speech is clear even at high recording speeds but is not as natural or smooth as larger neural network synthesizers, which are based on human speech recordings.	Preferred for environments requiring worldwide language coverage that are not sensitive to slightly 'robotic' accents
MaryTTS	MARY Text-to-Speech is a text-to-speech synthesis framework written in Java and supporting a broad spectrum of languages, including English (UK and US), French, German, Italian, Swedish, Telugu, Turkish, and Russian.	Optimal for multiplatform implementations as this framework is written entirely in Java

Continuation of the Table 3

1	2	3
Festvox/flite	Known as Festival Lite or Flite, Carnegie Mellon University's Flite is a speech synthesis engine written in C. It covers only English but is recognized for its rapid text-to-speech conversion, positioning it as a preferred solution for applications running on embedded devices.	Optimal for computational resource-sensitive speech generation contexts, such as embedded devices, where multilanguage support is not critical
Coqui-ai's TTS	Coqui-ai's TTS is a deep learning toolkit written in Python that is highly rated by the GitHub community for supporting fast and efficient model training for voice cloning in 16 languages. It optionally includes Coqui-ai's Open-Speech-Corpora library that provides more than 20 thousand hours of recorded audio in multiple languages available for fine-tuning translations and improving synthesized voice realism.	Applicable in projects where voice cloning models may need extensive training on large audio datasets to fine-tune voice realism
Tortoise TTS	Tortoise TTS is a multi-voice text-to-speech system written in Python and trained with an emphasis on voice quality. It allows training voice cloning models in different languages, but it doesn't train its models fast, and it is called a 'tortoise' for that reason.	Favored in speech generation environments where high speed of model training is not critical
PaddleSpeech	PaddleSpeech is a voice synthesizer toolkit written in Python and C++, currently covering only English and Chinese. It provides a comprehensive suite of speech technologies, including speech recognition, text-to-speech, voice pattern extraction, and speech translation.	Applicable in advanced speech processing and translation projects for English and Chinese in environments that are not sensitive to slightly 'robotic' accents
BenAAndrew's Voice-Cloning	The Voice-Cloning framework is written in Python and JavaScript and allows training voice cloning models in different languages. It uses a reworked version of the Tacotron2 framework and, therefore, discloses that all copyrights belong to NVIDIA and follow the requirements of a BSD-3 license.	Favored in projects where voice cloning models may need extensive training in different languages to fine-tune voice similarity and realism

Note: systematized by the author

Prospects for the development of audio deepfakes and their future. The prospects for the development of audio deepfakes are enormous. They are mostly related to using artificial intelligence (Hays, 2023). Let's consider the future of these technologies.

1. Audio Watermarking as a Security Enhancement. Audio watermarking emerges as a promising application, transforming the narrative from combating deepfakes to embedding authenticity within genuine audio content. Companies like Resemble.ai and Microsoft are adopting audio watermarking in their text-to-speech (TTS) products to show how security measures can be seamlessly integrated into the technology. This not only aids in the clear differentiation between authentic and deepfake content but also upholds the integrity of digital audio communications, ensuring that genuine content remains verifiable and trustworthy.

2. Benevolent Applications of Audio Deepfakes. The benevolent applications of audio deepfakes are diverse, covering entertainment, education, personalized content creation, and even therapeutic uses. For instance, generating hyper-realistic audio content can enhance audiobooks, making them more engaging by using the voices of historical figures or fictional characters. In educational settings, audio deepfakes can facilitate immersive language learning experiences or simulate historical speeches for a more interactive learning environment. Moreover, in personalized media, audio deepfakes can tailor content to individual preferences, creating unique and engaging user experiences.

3. Integration with Other AI Technologies. Integrating audio deepfakes with other AI-driven technologies, such as video deepfakes and virtual reality, opens up new opportunities for multimedia content creation. This synergy can lead to more immersive and interactive experiences in virtual environments, gaming, and online platforms. The capability to create cohesive audio-visual deepfake content can transform storytelling, entertainment, and even social interactions in digital spaces, offering a richer, more engaging user experience.

4. Advancements in Open-Source AI Technologies. Open-source platforms serve as catalysts for rapid advancements in AI, including audio deepfakes. These platforms not only accelerate the development and refinement of deepfake technologies but also empower a diverse community of developers, researchers, and enthusiasts to contribute to the evolution of robust detection mechanisms, such as audio watermarking. This collaborative environment turns the potential challenge of deepfake detection into a continuous cycle of innovation, where each contribution enhances the system's resilience against misuse (Hays, 2023; Fauve, 2023).

In the contemporary landscape, where cutting-edge technologies emerge at an unprecedented pace, the domain of audio deepfakes, particularly those powered by artificial intelligence, has gained significant traction. Among the emerging open-source projects in this field, Wunjo AI (Habr, 2023) stands out as a notable example that encapsulates the potential and versatility of AI-driven audio and video deepfake creation.

Wunjo AI is accessible through its GitHub repository, where users can find the source code, documentation, and installation instructions for Linux, MacOS, and Windows. The project aims to foster a community of contributors to develop further and refine its capabilities. Wunjo AI represents a step forward in making deepfake technology and speech synthesis more accessible and customizable, inviting a broad audience to contribute to and benefit from the project's development.

The main functionality of this program includes:

Text-to-Speech Synthesis. Converts written text into realistic speech using advanced NLP techniques. Offers three voice models (female, male, and robotic) in English, Chinese, and Russian, with support for custom Tacotron 2 voice models and phoneme format for English.

Deepfake Video Creation. Transforms images into videos by applying facial expressions and gestures, allowing for dynamic character animations. An extension enables image generation for deepfake videos using Dall-e 2, provided the images clearly show eyes and mouth.

Custom Extensions. Users can enhance Wunjo AI's functionality by creating extensions for various purposes, such as console interaction, GPU usage, voice model training, and ChatGPT integration (Habr, 2023).

Challenges and threats posed by audio deepfakes. While voice cloning and audio deepfake technologies hold significant promise for a range of professional applications, from automated news reporting to personalized content creation, they are beset with challenges that span ethical considerations, computational demands, linguistic diversity, and the intricacies of human speech (Khanjani et al., 2023).

1. Ethical Concerns and Misuse. One of the most pressing concerns with TTS and audio deepfake technologies is their potential for misuse. Products like Lyrebird's Speechify and Descript have demonstrated the ability to generate highly realistic speech rapidly, raising concerns about creat-

ing fake personas or fabricating audio for malicious purposes, such as spreading misinformation or impersonating individuals to stir political or societal unrest.

2. Computational Requirements. Creating synthetic speech is resource-intensive, requiring significant processing power and data storage. Although advancements in software efficiency have mitigated this issue to some extent, the creation and refinement of high-quality speech synthesis still demand substantial computational resources.

3. Dependence on Speech Corpus Quality. The quality of a TTS system is heavily reliant on the speech corpus from which it learns. Creating a comprehensive and high-quality speech corpus is not only expensive but also challenging, especially for languages or dialects with limited available data. Updating or modifying an existing speech corpus can be complex and resource-intensive.

4. Language and Dialect Limitations. TTS technologies face difficulties in accommodating sparsely spoken languages, particularly those without a standardized writing system. Additionally, the lack of linguistic components readily available for all languages poses a challenge for creating universally effective speech synthesizers. This limitation extends to accurately modeling dialect variations and accents, which is essential for producing speech that reflects genuine human language diversity.

5. Prosodic Challenges. Prosody, encompassing speech's rhythm, stress, and intonation, is crucial for the authenticity and intelligibility of synthesized speech. However, many TTS systems need help to implement prosodic features effectively, resulting in speech that can lack the nuanced emotional and phonetic qualities of natural human speech. The inability to accurately mimic human prosody can lead to synthesized speech that sounds unnatural or robotic.

6. Homographs and Special Characters. TTS systems often struggle with homographs–words spelled the same but with different meanings–leading to incorrect pronunciations in context. Furthermore, recognizing periods, special characters, and the nuances of human speech elements like breathing, laughter, and pauses remains challenging, deviating from synthesized speech's human-like quality (Khanjani et al., 2023).

The realm of deepfake detection in audio recordings grapples with formidable challenges, as the technology behind deepfake generation continues to evolve and become more sophisticated. At the forefront of this battle are advanced deep learning models and innovative datasets designed to train these models to discern the subtle discrepancies that differentiate genuine human speech from its artificially generated counterparts. Among the notable endeavors to enhance deepfake detection capabilities are extensive datasets like DFDC, DeeperForensics-1.0, and Celeb-DF, each offering a wealth of data aimed at capturing a broad spectrum of human speech characteristics across different genders, ages, and ethnicities. These datasets are instrumental in training detection algorithms to recognize and flag inconsistencies and anomalies in audio samples that may suggest manipulation. However, despite these advancements, the detection of audio deepfakes remains fraught with difficulties. The technology's capacity to generate increasingly convincing fake audio poses a persistent challenge, necessitating ongoing research and development to refine detection methods and ensure they remain effective against an ever-evolving threat landscape (Naitali et al., 2023).

Conclusions. Audio deepfakes represent a significant technological advancement in the creation and manipulation of audio recordings to mimic individual voices convincingly. The three primary categories of audio forgeries – imitation-based, synthetic-based, and voice cloning – each serves different purposes and presents unique challenges. While these technologies offer innovative applications in entertainment, customer service, and security testing, they also raise ethical and security concerns, necessitating careful consideration and regulation to mitigate potential misuse.

The landscape of audio deepfake creation is populated by a diverse array of software solutions, each tailored to specific needs and applications. Notable programs include Resemble.ai for entertainment, Descript for podcasting and audio editing, CereProc for multilingual voice cloning, Respeecher for filmmaking and gaming, iSpeech for customer service and gaming applications, ReadSpeaker.ai

for business narration, Replica Studios for video games and voiceover projects, and Voice Changer for fun and entertainment in mobile applications. These tools, alongside open-source technologies like Wunjo AI, eSpeak NG, and CorentinJ's Real-Time-Voice-Cloning, provide a comprehensive toolkit for professionals and enthusiasts to explore the creative and functional potentials of audio deepfakes, with a strong emphasis on ethical use and innovation. The future of audio deepfakes presents both exciting opportunities and significant challenges. The technology holds great promise for innovation in storytelling, personalized content, and educational tools, supported by advancements in AI and open-source platforms like Wunjo AI. However, ethical concerns, computational demands, and linguistic diversity pose considerable challenges. The potential for misuse in spreading misinformation and the technical complexities involved in creating realistic and diverse linguistic content highlights the need for ethical guidelines and robust detection mechanisms. Balancing the benefits against the risks will be the key topic in the development and application of audio deepfakes in the next several years.

References:

1. Almutairi, Z. & Elgibreen, H. (2022). A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions. Algorithms. 15. 19. 10.3390/a15050155. URL: https://www.researchgate.net/publication/360354997_A_Review_of_Modern_Audio_Deepfake_Detection_Methods_Challenges_and_Future_Directions
2. Broz, M. (2023). Top 10 Deepfake Voice Software & Online Tools Review. VansMedia. URL: https://vansmedia.vanceai.com/deepfake-voice-software-and-online-tools-review/
3. Fauve, B. (2023). The Rise of Audio Deepfakes: Implications and Challenges. URL: https://www.linkedin.com/pulse/rise-audio-deepfakes-implications-challenges-benoit-fauve
4. Habr (2023). Создание deepfake видео и синтез речи open-source проект Wunjo AI. URL: https://habr.com/ru/articles/752910/
5. Hays, E. (2023). Beyond the Horizon: Exploring Future Prospects and Societal Impacts of Deepfake Technology. URL: https://medium.com/@toddkslater/beyond-the-horizon-exploring-future-prospects-and-societal-impacts-of-deepfake-technology-ecc53b51fcb2
6. Kamunya T. (2023). 8 Best Open Source Deepfake Software for Realistic Illusions. URL: https://geekflare.com/best-open-source-deepfake-software/
7. Khanjani Z., Watson G., Janeja V. (2023) Audio deepfakes: A survey. Frontiers, 5. URL: https://www.frontiersin.org/articles/10.3389/fdata.2022.1001063/full
8. Kietzmann, J., Mills, A. & Plangger, K. (2020). Deepfakes: perspectives on the future "reality" of advertising and branding. International Journal of Advertising. 40. 1–13. DOI: 10.1080/02650487.2020.1834211.
9. Naitali, A, Ridouani, M, Salahdine, F, Kaabouch, N. (2023) Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions. Computers, 12(10):216. DOI: https://doi.org/10.3390/computers12100216
10. Whittaker L., Mulcahy R., Letheren K. (2023). Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. Technovation, Volume 125. URL: https://doi.org/10.1016/j.technovation.2023.102784