UDC 004.8:004.056

DEEPFAKE TECHNOLOGIES IN IT-SPHERE

Deikalo I.S, Volfovich V.D

Belarussian State University of Informatics and Radioelectronics, Minsk, Belarus

Perevyshko A.I – Senior Lecturer at the Department of Foreign Languages

Annotation. This article is about Artificial Intelegence/Machine learning deepfakes in present world. At the beginning the origin of a term Deepfake are given. After that types of deepfakes are categorized. At the end recommendations for avoiding Deepfake scam are provided.

Keywords: Deepfake, audio, video, image, fraud, AI

Introduction. Deepfake («deep learning» and «fake») is a term which refers to artificially generated media created using advanced AI technologies of this category. These tools are used to manipulate images, videos, and audio content, eventually blurring the line between what is real and what is not. The risks that have come up with deepfakes are wide, touching a lot of areas. The more alarming dangers are related to the spread of misinformation and how they manipulate the public. For example, a video showing a public figure making false statements could be used to influence elections, cause violent outbreaks, or even tarnish a reputation [1]. Deepfakes may also find more use in cybercrime, now that there may be identity theft, financial frauds, and blackmail. This involves sharing sensitive information or transferring money when impersonating someone [2].

Main part. Deepfakes are one of the products of advanced Artificial Intelligence and Machine Learning technologies. They are classified into three major types based on the medium serving as the manipulated channel: Video, Audio, and Image. The precondition-underlying techniques apply differently and pose different challenges and risks respectively to each. Video deepfakes are movies where the face or body of one person is replaced with that of another, who will then be shown doing or saying things they never did or said. Greatly realistic results are due to the use of techniques like generative adversarial networks (GANs) that involve facial expressions, movement, and lighting (shown in Figure 1).

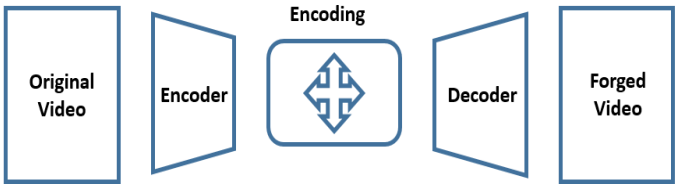

Figure 1 – Deepfake video creation process

In entertainment, video deepfake technology may develop as a basis for footage in memes, parodies, and films with digital clones of deceased actors. It may also be utilized as a channel to assault social trust through the fabrication of videos depicting public figures. Though such applications may seem frivolous, deepfake technology remains extremely dangerous [3]. It can help proliferate fake news and manipulate public opinion by eroding trust in the video as a respectable source of information. It may also be used in creating sexual content without the depicted person's approval. Audio deepfake technology, alternatively known as voice cloning, identifies a replicated speech synthesis from AI technology with coherent speech from the victim. What it really does is break the speech of a victim down into minimal phonetic units, which are

then concatenated to result in humanly distinguishable speech [4]. These systems extract vocal patterns, intonation, and emotional nuances to construct speech that is nearly indistinguishable from the original (shown in Figure 2).

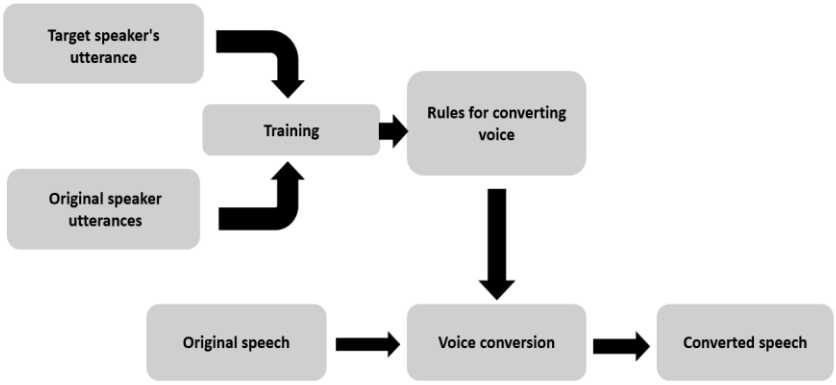

Figure 2 – Deepfake audio creation process

Much-improved quality in audio deepfakes has been experienced with more sophisticated models like WaveNet and Tacotron. Commercial applications of audio deepfake technology range from the development of personalized voices for virtual assistants and audio-visual content to mere entertainment. The technology can also be used for malicious purposes, such as fraud where an individual can impersonate someone else in phone conversations or voice messages to gain some form of benefit. Image deepfake concerns the creation or manipulation of images using artificial intelligence technology. Examples are face swapping, the creation of fully synthetic images, or editing certain fragments of a photo, such as inserting or removing objects. More commonly, imaginations are done using generative adversarial networks to come up with visually appealing images that would pass close inspection on their authenticity. The applications for image deepfakes are very wide, ranging from artistic creation and design to social media. They can get applied for developing ad visuals or digital artwork. On the down part, they work their way on risks such as creating false evidence (staged crime scenes) and for cyberbullying, cancel culture, and blackmail [5].

Conclusion. How not to get involved in fraud created by deepfakes? Stay aware and notice everything. Try not to make quick judgments or decisions under pressure and overemotions. Check the videos for minor things, like if there are any unnatural moments of facial expressions, repeating mechanical speech, and inconsistent backgrounds. If there's an offer that's not normal, like needing to give immediate financials or an appeal made from emotions, try to contact the person through a known different channel- direct telephone call or a physical meeting. Avoid sharing your passwords, verification or bank information if someone asks for them through unsolicited messages even if they appear genuine, personal data should never be given away here. Use two-factor authentication and constant software updates. Knowledge about scams can be known through, e.g, CISA. And any suspicious activities must be reported to the concerned authorities. Build a trusted capacity of good skepticism for careful investigations and start the urgency of sharing details over AI inflow of attacks.

References

1. *Chesney B, Citron D (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. Calif L Rev 107(6):1753–1820.*

2. *Hancock JT, Bailenson JN (2021) The social impact of deepfakes, cyberpsychology, behavior, and social networking. Mary Ann Liebert Inc., pp. 149–152.*

3. *A. Firc, K. Malinka, P. Hanáček (2023) Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors. Heliyon,p. e15090.*

4. *Albahar M., & Almalki J. (2019). Deepfakes: Threats and countermeasures systematic review. Journal of Theoretical and Applied Information Technology, 97(22), 3242–3250.*