EXPLORING THE DEEPFAKE DILEMMA

Mihaela UNTU*, Victoria MUTRUC, Daria RAȚEEVA

Department of Software Engineering and Automation, group FAF-232, Faculty of Computers, Informatics and Microelectronics, Technical University of Moldova, Chișinău, Republic of Moldova

*Corresponding author: Mihaela UNTU, mihaela.untu@isa.utm.md

Tutor/coordinator: **Elena GOGOI**, university lecturer, Department of Software Engineering and Automation, Faculty of Computers, Informatics and Microelectronics, Technical University of Moldova, Chișinău, Republic of Moldova

Abstract. *Nowadays, the emergence of deepfake technology implies major ethical concerns. At this stage of artificial intelligence advancement, there is a thin line between real and fabricated content, raising profound concerns regarding the ethical implications of its use and creation. The purpose of this article is to weigh on the jeopardy of deepfakes and the ethical concerns around it, in order to provide a better understanding of how artificial intelligence content influences our lives. The paper investigates the deepfake subject starting from understanding the core significance of what a deepfake is and ending with the future of this revolutionary technology. In addition, it provides tips on how to detect falsified content, such as manipulated photos and videos, to ensure safety online. Moreover, it explores the necessity of extending the legal framework regarding solutions against blackmail, intimidation, sabotage and scams. Finally, the paper presents ways in which threats can be turned into opportunities.*

Keywords: *audio and visual content, artificial intelligence, scams, technology*

Introduction

Nowadays, technology advancement is unavoidable and so the Artificial Intelligence field has undergone plenty of changes that has led to the development of deepfakes. Not only is the industry developing fast, but this type of technology is also becoming more and more accessible to the public. According to sumsob.com, there has been recorded a 10 times increase of deepfake content from 2022 to 2023 [1]. Deepfakes are known to be a contributing factor to the serious problem of "fake news" since they enable the mass production or alteration of media with the intention of disseminating false information. The difficulties presented by deepfake technology and the changing legal and protection landscape are discussed in this article, which also provides helpful advice. A few crucial questions need to be addressed in light of the concerning increase in deepfake content: What laws have been passed or are set to be passed, how can people protect themselves and their loved ones, and how can they identify deepfakes?

The rise of deepfakes

Living in a digital world, modern issues are inevitable to be avoided. Among these challenges, the deepfake technology has expanded and obtained attention for its ability to manipulate and exploit. According to the Oxford Dictionary, a deepfake refers to videos in which a person's face or body has been digitally altered to resemble someone else, often with malicious intent or to disseminate false information [2]. This term includes the malicious content which is produced without the consent of the individuals participating in it. Artificial Intelligence (AI) plays a core role in the growth of deepfake technology, expanding diverse realms from simple systems to complex audio and visual manipulation through deep generative processes.

A major concern of deepfake technology is the continuous integration of individuals into visual content without their consent, i. e., any person, anywhere in the world, can unconsciously become part of fabricated content which lately is spreading misinformation and jeopardizing trust in media. While fabricating deceitful content is not a new phenomenon, deepfakes bias advanced machine learning algorithms used to create audio and visual material that can persuasively capture unsuspecting viewers. This kind of technology has the potential to fabricate evidence of events that never occurred, therefore complicating the truth and authenticity of actions.

The article's author, Stu Sjouwerman, founder and CEO, KnowBe4, argues that while not a recent development, deepfakes have become a notable concern due to advances in machine learning and artificial intelligence. These technological advances allow cybercriminals to create remarkably convincing counterfeit audio and video. These targeted attacks have proven effective, causing hackers to refine their methods for greater profitability. The prevalence of deepfake videos increased by 84% between December 2018 and October 2019, likely underestimating the true extent of their proliferation. While many of these videos feature adult content, their potential for harm extends far beyond that, especially when considering the potential repercussions on business reputation and functionality. Without disclosing specific cases, Symantec documented three successful deepfake audio scams that tricked three CFOs into transferring substantial amounts of money. Forrester estimated that fake scams cost companies $250 million in 2020 [3].

As time goes by, it becomes harder and harder to distinguish deepfakes from real footage and prevent misinformation. Despite that, there are a couple of things one can look out for, in order to recognize deepfakes. According to media.mit.edu, when it comes to videos, your attention should be first drawn to the person’s face, because deepfake videos mainly revolve around facial transformations. An important difference between a video of a real person and a deepfake is an unnatural eye movement. This can be characterized by lack of or too much blinking. Another important factor in detecting AI generated content is represented by the mouth movements. They might look unnatural or out of sync with the audio. Other signs one should look for are: skin that is too smooth or too wrinkly, weird or lack of facial hair and lack of shadows.

According to Euro News, for photos, one should first look for odd details like unrealistic lighting, disfigured hands or picture-like graphics. Another important sign of fake contents is unnatural skin tone with skin that is too smooth. Moreover, pictures can feature unnecessarily blurred details and incorrect writing of words. The background can also suggest that the picture is fake if it features abnormal looking objects or if the background does not correspond to the city/location where the photo was supposedly taken [4].

The ethical dilemma of deepfakes

Even though deepfakes show how far humanity has come when it comes to Artificial Intelligence and they can be used for positive things, their apparition still has raised multiple ethical concerns, like consent, privacy and responsability. AI content can lead to mass misinformation, defamation cases and manipulation of public opinion. Because deepfakes can use a real person’s face, voice and body without their consent, individuals can find themselves in compromising situations and therefore get their reputation damaged. Moreover, deepfakes can be used to potentially scam other people or even businesses. No one can ensure that the people that do get to use the technology will use it responsibly and ethically. This is why deepfakes are seen as a threat to people’s right to privacy. Since the development of deepfake technology, there have been numerous cases where it was utilized in illegal activities, leading to significant scandals. In some cases this activities caused major problems. Two notable examples include:

First example of a serious scandal of such type was the CEO Fraud using Deepfake Audio: An unusual case involved the use of deepfake audio, an AI-generated audio, in a CEO fraud

scheme that reportedly swindled US$243,000 from a U.K.-based energy company. According to "Cyber Attacks" blog, fraudsters employed voice-generating AI software to replicate the voice of the chief executive of the company's Germany-based parent company. This deception was utilized to facilitate an unauthorized fund transfer [5].

Another case when deepfake lead to a serious scandal, according to Heather Chen and Kathleen Magramo, CNN, in their article "Finance worker pays out $25 million after video call with deepfake 'chief financial officer'" was the case when a $25.6 Million Theft from a Multinational Finance Firm was stolen. Scammers utilized deepfake technology to steal $25.6 million USD (equivalent to $200 million Hong Kong dollars) from a multinational finance firm. Hong Kong police were alerted to the situation, which involved scammers creating a deepfake video impersonating the firm's chief financial officer (CFO) in a video call. During the call, the scammers interacted with an employee and other company staff members, all of whom, except for the employee, were deepfake replicas. The fake CFO instructed the employee to carry out 15 separate financial transfers totaling $25.6 million USD. The scam was only discovered a week later when the employee realized the deception after speaking with colleagues [6].

Legal framework of deepfakes. A look at global deepfake regulations

"From its nascent development in the 1990s to the introduction of a widely available app in 2018, deepfake technology has become both increasingly sophisticated and readily accessible to the general population" [7]. According to the Princeton Legal Journal, there is currently no federal legislation addressing the prospective threats of deepfake technology in the United States. However, in December of 2019, Congress passed the National Defense Authorization Act (NDAA), which, in Section 5709, requires the Director of National Intelligence to report on the use of deepfakes by international governments, its ability to spread misinformation, and its potential impact on national security [8]. However, several US states including Georgia, Florida, Hawaii, Tennessee, New York, and others, have implemented deepfake laws, or are still in the process of implementing. They vary depending on the state. Meanwhile, the European Union has already agreed on the world's first comprehensive AI law. AI Act will be responsible for this legislation and regulate deepfakes around the EU [9].

Unraveling the opportunities and consequences of deepfakes

With all these scandals and prejudice caused by deep fakes, the question appears, what will happen in the future? Will deepfakes be banned or will it remain a source of scandals and scams, will any laws appear around use of deepfakes and in the end will deepfakes be used for noble purposes? According to Sudhanshu Kumar in his article "The Future of Deep Fakes in the World of Democratized Artificial Intelligence", in the near future, as technologies like machine learning and deep learning continue to advance, we could expect a significant shift toward the democratization of artificial intelligence. This transition will involve intelligent algorithms taking over many manual procedures and processes, fundamentally changing the way data is collected and managed. As these algorithms gain access to more refined and comprehensive data, their capabilities will evolve, leading to increased efficiency and effectiveness. Consequently, fraud detection mechanisms will become more accurate and less prone to ambiguity. Advanced artificial intelligence systems will be able to identify various forms of fraud, including deep fakes, thus revolutionizing current paradigms in fraud detection. These systems will not only elucidate the methods used in fraud detection, but also provide evidence-based explanations for their findings. Moreover, they will provide proactive recommendations for preventive measures [10].

Even though deepfakes can cause great damage when used irresponsibly, the further development of this type of technology could potentially be used for a good cause. Deepfakes

contribute to the creation of new forms of digital entertainment. This type of technology can serve as a dubbing tool, so the facial expressions and voice of the person speaking can appear as if the person is actually speaking another language. This can be used not only for dubbing movies and ads but also for translating educational content, like lectures. It can also be used in film making for more affordable production. Moreover, if one of the actors dies or ages to the point that they do not resemble the character anymore, deepfake technology might be used to bring the character back to life in the movie. AI technology can contribute to the development of the educational field as well. It is possible to make historical classes more interesting and appealing to students by bringing historical figures back to life [11].

Conclusions

The article depicts the risks associated with deepfake technology, highlighting its ability to create persuasively fabricated content that undermines trust in media and spreads misinformation. The ways in which one can protect oneself from mistaking deepfake content for real one are also mentioned.

Moreover, it describes the development of deepfake technology, its legal framework challenges in the US and the more advanced legislative response of the European Union. The article later presents the major scandals produced by the use of deepfakes. These burglaries can lead to serious consequences for the victims and enormous losses for the companies that were involved in that scandal. Even though these days there are a lot of existing and potential threats regarding content generated by Artificial Intelligence, it is important to understand that there are some opportunities that come with this kind of technology and it is expected of people to use these resources responsibly.

References

[1] "Sumsub Research: Global Deepfake Incidents Surge Tenfold from 2022 to 2023." *Sumsub*, 28 November 2023, [Online]. Available: https://sumsub.com/newsroom/sumsub-research-global-deepfake-incidents-surge-tenfold-from-2022-to-2023/.

[2] S. Cole, "deepfake, n. meanings, etymology and more." *Oxford English Dictionary*, [Online]. Available: https://www.oed.com/dictionary/deepfake_n?tl=true.

[3] S. Sjouwerman, "The evolution of deepfakes: Fighting the next big threat." *TechBeacon*, [Online]. Available: https://techbeacon.com/security/evolution-deepfakes-fighting-next-big-threat.

[4] I. El Atillah, "How to spot a deepfake: 5 things to watch out for to identify AI-generated content online", Euro News, last modified 09/05/2023, [Online]. Available: https://www.euronews.com/next/2023/05/01/how-to-spot-an-ai-deepfake-midjourney-from-trump-arrest-to-the-pope-puffer-coat.

[5] "Unusual CEO Fraud via Deepfake Audio Steals US$243,000 From UK Company - Noticias de seguridad." *Trend Micro*, [Online]. Available: https://www.trendmicro.com/vinfo/mx/security/news/cyber-attacks/unusual-ceo-fraud-via-deepfake-audio-steals-us-243-000-from-u-k-company.

[6] S. Kearns and A. Sacal, "Scammers in Hong Kong Used Deepfakes To Steal $25.6M USD." *Hypebeast*, 5 February 2024, [Online]. Available: https://hypebeast.com/2024/2/scammers-hong-kong-deepfake-technology-theft.

[7] "The History of Deepfake Technology: How Did Deepfakes Start?," Deepfake Now, last modified April 21 2020, [Online]. Available: https://deepfakenow.com/history-deepfake-technology-how-deepfakes-started/.

[8] "The High Stakes of Deepfakes: The Growing Necessity of Federal Legislation to Regulate This Rapidly Evolving Technology - Princeton Legal Journal." *Princeton Legal Journal*, 19 June 2023, [Online]. Available: https://legaljournal.princeton.edu/the-high-stakes-of-deepfakes-the-growing-necessity-of-federal-legislation-to-regulate-this-rapidly-evolving-technology/#_ftn1.

[9] A. Owen, "Deepfake laws: is AI outpacing legislation?" *Onfido*, 2 February 2024, [Online]. Available: https://onfido.com/blog/deepfake-law/.

[10] "The Future of Deep Fakes in the World of Democratized Artificial Intelligence." *Diplomacy & Beyond Plus*, 14 January 2023, [Online]. Available: https://diplomacybeyond.com/the-future-of-deep-fakes-in-the-world-of-democratized-artificial-intelligence/.

[11] M. Kalmykov, "Deepfake Technology in Video Industry." *DataArt*, 28 November 2023, [Online]. Available: https://www.dataart.com/blog/positive-applications-for-deepfake-technology-by-max-kalmykov.