AGORA International Journal of Juridical Sciences, http://univagora.ro/jour/index.php/aijjs
ISSN 1843-570X, E-ISSN 2067-7677
Vol. 18, No. 2 (2024), pp. 266-271

DEEP FAKE VULNERABILITIES IN PUBLIC ADMINISTRATION DURING THE ELECTIONS PERIOD

M. A. VANCU

Mircea Alexandru Vancu
”1 Decembrie 1918” University of Alba Iulia, Romania
E-mail: vancualexandru90@gmail.com

Abstract: *The speed of technology development, digitization, has marked a radical transformation of our society, managing to significantly influence the way we communicate, inform ourselves or interact. The technological advance in the field of artificial intelligence has created the conditions for the emergence of a new form of manipulation that is still widespread in the online environment: deepfakes. Deepfakes become a problem when fake content can be created to harm a private person's reputation or life. In situations regarding future elections and candidate selection, a malicious deepfake message to put the adversary in a position of incompatibility for election to public office or to manipulate public opinion regarding the possibility of certain candidates being "suitable" for a certain function, knowing this type of tool can mean the difference between winning or not.*

Keywords: *Deepfake technology, Social engineering, Public administration, Manipulating elections through deepfake technology*

1. Introduction

Public life has become a particularly important asset, the image is well formed, structured and delimited over time and yet so vulnerable to the problems that have arisen with digitalization. In the National Strategy for Artificial Intelligence (AI) 2024-2027, we recognize that the use of AI technology makes it possible to be attacked by a new type of threat, hybrid warfare, through disinformation and influence operations. While altering images or video is not as much of an issue, large-scale access to deep fake technology can be a threat. Technological advances in artificial intelligence and digitization have made it possible to create a new form of digital manipulation, the deepfake. Deepfake is a form of digital manipulation that uses advanced artificial intelligence techniques to create fake images or audio-video content. It can have significant consequences for society, including impacts on security and public trust in online information. In this context, it is crucial to be aware of the new challenges in this area and therefore to be able to identify and manage fake content appearing online. Public life has become a particularly important asset, the image is well formed, structured and delimited over time and yet so vulnerable to the problems that have arisen with digitalisation. In the National Strategy for Artificial Intelligence (AI) 2024-2027, we recognise that the use of AI technology makes it possible to be attacked by a new type of threat, hybrid warfare, through disinformation and influence operations. While altering images or video is not as much of an issue, large-scale access to deep fake technology can be a threat. Technological advances in artificial intelligence and digitisation have made it possible to create a new form of digital manipulation, the deepfake. Deepfake is a form of digital manipulation that uses advanced artificial intelligence techniques to create fake images or audio-video content. It can have significant consequences for society, including impacts on security and public trust in online information. In this context, it is crucial to be aware of the new challenges in this area and therefore to be able to identify and manage fake content appearing online.

The rapid development of technology and communications has brought great benefits to everyone, and it is only logical to understand that this has also increased vulnerability to cyber-attacks. Increasing digitization, connectivity as well as the use of new technologies have led to worsening cybersecurity risks. To this end, the National Cyber Security Directorate was established by GEO no. 104/2021. The National Cyber Security Directorate is a Romanian institution under the coordination of the Prime Minister. DNSC's main tasks are prevention, analysis and response to cyber incidents, cyber awareness and international cooperation. This institution has a particularly important role as cyber security is a major concern in today's digital society. All national critical infrastructures such as power grids, financial systems and health systems can be vulnerable to cyber-attacks. DNSC has an important role in protecting these infrastructures and ensuring Romania's cybersecurity.

The Directorate of National Cyber Security (DNSC) released a guide to identify deepfake material in a press release on the first day of April 2024. The guide is a well-developed guide that provides detailed information on what deepfake is, the process of how it is made and how to identify such material. By understanding these concepts, users will become more aware of the risks associated with this type of technology and gain knowledge on how to defend themselves. Understanding the concept of deepfake is crucial and has significant implications for society, politics, cybersecurity and the manipulation of public trust. The deepfake guide is a trusted source of information that explains how deepfakes can distort reality through video or audio content that is difficult to identify from authentic material and how this can undermine trust.

One role of the guide may be to better understand how we rely on information to make decisions and of course to use more critical thinking and to check the veracity of information and its sources. This paradigm shift in which access to information is close to 100% compared to times in the past when access was much harder and information was much harder to find, makes us vulnerable to novelty. We can observe the adverse effects of this era, namely the fact that nowadays we can find a plethora of information results requested through the internet, but we lack a reliable source of its veracity. In this respect, we come up against information from false registrations involving people in activities that they have not carried out, false identities through which unauthorized access to information or financial resources is sought.

In public administration we can see the strongest impact because this is the most suitable area for attacks to manipulate and misinform public opinion. It is within anyone's reach to use a tool that can discredit the adversary and create confusion, and even more if it achieves its purpose, to influence important decisions such as influencing elections or conspiracy theories. This type of technology has a strong social, ethical and moral impact and without the right information and decisions can become a threat to anyone. In order to defend ourselves against this type of attack, we need to prepare for the future and understand how this technology works. At a larger level, there are technologies that are able to thwart attacks of this kind and that build on exactly what makes this technology so special.

The use of artificial intelligence and machine learning algorithms is a great way of using deepfake technology and that is a perfectly calibrated surgical tool for this type of activity. We therefore understand that it is important that we develop a collective level of understanding of deepfake and its destructive potential. We are in the digital age, a world in which technology is the greatest challenge to the human factor because it is increasingly difficult for us to distinguish reality through the manipulation of content. However, there are certain clues that can betray deepfake material. The only constant is change and understanding phenomena that create effects for everyone.

In order to understand even more and to speak from the experience of what we have learned, we can take as an example the communication through social networks in the coronavirus pandemic. Fake news was a very effective way of manipulation and disinformation. Cifuentes-Faura (2020) explains how people react in an emergency or disaster. The rapid lack of a prompt response from the authorities makes citizens seek information themselves in the media or through social networks but without verifying the source of this information. In this way misinformation and fake news create panic and promote inappropriate attitudes and behavior. One of the main distribution channels for fake content is WhatsApp. The app is the most suitable channel for spreading fake news because it works based on direct trust in the people we have in our address book, in the sense that people share what they get from people in their environment and much of the information shared is false, managing to lead to wrong decisions. We are therefore identifying the need for a reference guide to check the veracity of information so that we do not find ourselves in the situation of being yet another source of false news. The most important and first signal that we might be dealing with fake news is the headline. In the age of social media, the headline needs to be short, concise and visual. In the case of fake news, headlines are often too prominent and include information that is hard to believe. In the top three most searched headlines on the Google search engine in 2020 was the coronavirus pandemic but also a movie that referred to a similar pandemic. I wanted to highlight the fact that by understanding the trends, you can easily produce material that for some people becomes a reality and that they no longer question the aspects presented or the fact that it is not real. In this sent we can deduce that the non-identification of the owner of the news, the information of the communication channel and the fact that it uses trends is another sign of fake news.

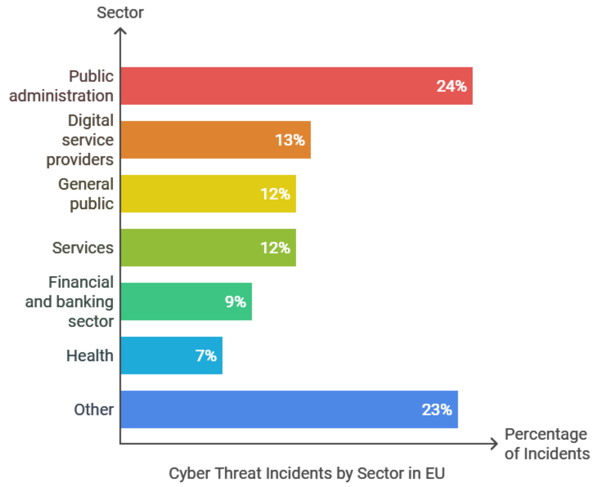
The fake news from the coronavirus pandemic was usually unknown media sources offering exclusivity for certain channels. This is false given that the importance of the message of a news message that everyone is waiting for, such as the vaccine or how best to protect ourselves from coronavirus, is something that needs to be carried on all channels. On top of that, to raise even more questions, the format of the message is atypical and gives even more power and diversion. False information is usually posted without the date being given, and some have spelling mistakes and are found on media-like internet addresses. In order to provide an official news source, the Romanian Government, through the Romanian Digitization Authority and the non-governmental organization Code for Romania, launched the online platform for official news in 2020. The findings presented by Shweta et al. (2021) refer to the fact that the recent advancement in the field is not effective as long as there is so-called ,,gap" with the meaning of incomplete work . Citizens can use effective identification methods that detect deepfakes, but this requires certain knowledge in the field that not everyone has. So the authorities could step in and implement a solution like the Screening presented in the review above, and legal requirements could be made for tech companies to be obliged to include limitations in the platforms through which people create digital content or create metadata that is fingerprinted by time, location or content.

2. Manipulating and stealing public image

Social engineering is just one of the threats the European Union is tackling to promote cyber resilience and combat cybercrime. Social engineering is a strategy used by individuals or groups of individuals to manipulate and mislead people into disclosing sensitive information that compromises them or others to whom they have access (https://www.consilium.europa.eu/ro/policies/cybersecurity/cybersecurity-social-engineering/).

This is based predominantly on human psychology and understanding of human behavior. In social engineering attacks, the attacker often pretends to be a trustworthy person or a trusted source to gain the trust of the intended victim. Tricking victims into opening malicious documents, files or emails, or visiting certain websites thus giving the attackers unauthorized access to systems or services. The most common attack of this kind is phishing (via e-mail) or smishing (via text messages). The attacker will use tactics such as identity theft, means of persuasion to extract valuable information. In public administration social engineering attacks can take various forms, the strategy being to exploit human vulnerabilities to achieve objectives. Because we are talking about the central public administration system it goes without saying that we are dealing with a very well-prepared system if we are to talk about the strong, technical link in terms of data protection and security and cyber security. The weak link, on the other hand, can be considered to be man.

Figure 1. Looking at the top 6 sectors affected by cyber threats



Source: https://www.europarl.europa.eu/topics/ro/article/20220120STO21428/securitate-cibernetica-principalele-amenintari

The 24% percentage of reported incidents in public administration related to the main threats observed by the European Union Agency for Cyber Security in 2022 has given us an indication that civil servants may be at risk of security incidents.

Cyberspace is characterized by its border lessness, dynamism and anonymity, which creates both opportunities for the development of the knowledge-based information society and risks to its functionality, whether at the individual, state or even cross-border level. By this we mean the duality that the more computerized a society is, the more vulnerable it is, and that ensuring the security of cyberspace must be a concern for all state institutions. A good example to follow is the course of the Information Technology and Cybersecurity Service organized in Moldova and funded by the European Union, which organized in April 2024 a course for civil servants from various government institutions with the aim of providing participants with an understanding of cybersecurity concepts and practices to protect information assets in the online environment (https://stisc.gov.md/ro/comunicate-de-presa/stisc-continua-instruirea-iso-securitate-cibernetica-functionarilor-publici).

In order to compromise a person, we can identify different techniques used in social engineering such as phishing, pretexting, bait phishing, spear-phishing, vishing, ransomware and extortion, luring and identity theft.

- **Phishing**: Attackers send misleading emails, messages or links to legitimate-looking websites to persuade recipients to click and reveal sensitive information such as passwords, credit card numbers or personal data;

- **Pretexting**: Attackers create a fabricated scenario or pretext to obtain information. This often involves impersonating a trustworthy person such as a colleague or bank representative;

- **Bait Phishing**: Attackers mainly target users of social media platforms and often take advantage of popular topics or trending topics to create misleading messages, making them relevant and trustworthy;

- **Spear-phishing**: Similar to phishing, but highly targeted. To appear even more convincing, attackers tailor messages based on information about the potential victim;

- **Vishing**: A form of phishing that occurs through voice communications, usually phone calls. Attackers impersonate trusted identities and persuade victims to reveal sensitive information;

- **Ransomware and extortion**: Attackers threaten to disclose sensitive information or disrupt systems unless a ransom is paid. They instill fear to coerce their victims;

- **Luring**: Attackers offer something attractive, such as free software downloads, in exchange for personal information or system access. This may involve infected files or links;

- **Impersonation:** Attackers pretend to be someone else, either online or in person, in order to gain the victim's trust and manipulate them into divulging confidential information or performing certain actions.

To be able to avoid these types of attacks, we must be vigilant and recognize the signs. Something that seems too good to be true can always be a scam. Sensitive, institutional or personal data should not be shared, such as passwords, credit card numbers or personal data in messages or emails, regardless of who requests it. We should also avoid clicking on links or opening emails from unknown sources.

A highly publicized example of a deep fake scam was in Hong Kong Japan, where through social engineering, a single person was chosen as a target. The person was sent some emails to make some bank transfers. The Hong Kong fraud most likely used deep-fake technology in real time, the attacker managing to achieve his goal. To confirm the transfers requested by email, the victim was called and put in a conference call with some of the employees and the company's CEO, all the content being fake, to confirm that he had to make those requested transfers and that he did. The value of the fraud was over 25 million dollars. This is a good example of how deepfake technology and social engineering were used and which makes us think about the risks to which each of us is exposed and even more so a public institution.

Detecting a deepfake is quite difficult but a trained eye can still detect it. In the early days of deepfakes, detection was relatively easy and simple by observing whether the eyes blinked or not. Eyes that did not blink signaled a deepfake, because the technology was not advanced enough to create blinking eyes. Technology has advanced and this is more difficult to detect. Other signs that we can look at are (https://dnsc.ro/citeste/dnsc-scamadviser-deepfake-ce-este-recomandari-securitate-online): • Improper lip sync; • Unnatural blinking; • Stuttered movements; • Missing shadows and unnatural lighting; • Unmoving hair or unnatural hair movement.

Conclusions and proposals

Knowledge of deepfake technology can mean an extra step for those who will not fall into this manipulation trap, but it is not enough as we have observed. Actors interested in manipulating public opinion in favor of an event candidate in the elections can achieve their goal through social engineering. Public institutions are vulnerable to possible cyber-attacks, but especially to direct attacks on the person. However, by acquiring a minimum of information such as that in the guide posted in April 2024 on the website of the National Directorate of Cyber Security, they can form a filter through which reality is closer to the truth. This paradigm shift achieved through digitalization makes us understand how important it is to define a context for new technologies and how controlling them is necessary through a clear delimitation of competencies**.**

REFERENCES

1. Cifuentes-Faura, J. (2020). Fake news durante la COVID: ¿Cómo detectarlas?. Comunicación, (42), 100–103, https://doi.org/10.18566/comunica.n42.a07
2. Shweta Negi, Mydhili Jayachandran, Shikha Upadhyay (2021). *Deep fake: An Understanding of Fake Images and Videos*, International Journal of Scientific Research in Computer Science, Engineering and Information
3. Cybersecurity Strategy of Romania of 30th December 2021 for the 2022-2027 period [online] https://legislatie.just.ro/Public/DetaliiDocumentAfis/250235 [30.11.2024]
4. Emergency Ordinance No. 104/2021 on the establishment of the National Cybersecurity Directorate [online] https://legislatie.just.ro/Public/DetaliiDocumentAfis/246652 [30.11.2024]
5. https://dnsc.ro/citeste/dnsc-scamadviser-deepfake-ce-este-recomandari-securitate-online
6. https://stisc.gov.md/ro/comunicate-de-presa/stisc-continua-instruirea-iso-securitate-cibernetica-functionarilor-publici [30.11.2024]
7. https://www.consilium.europa.eu/ro/policies/cybersecurity/cybersecurity-social-engineering/ [30.11.2024]
8. https://www.europarl.europa.eu/topics/ro/article/20220120STO21428/securitate-cibernetica-principalele-amenintari [30.11.2024]