Blagoj Nenovski
University "St. Kliment Ohridski" – Bitola
e-mail: blagoj.nenovski@uklo.edu.mk
ORCID ID: 0000-0001-7093-5133

Ice Ilijevski
University "St. Kliment Ohridski" – Bitola
e-mail: iilijevski@uklo.edu.mk
ORCID ID: 0000-0002-2842-3633

Angelina Stanojoska
University "St. Kliment Ohridski" – Bitola
e-mail: angelina.stanojoska@uklo.edu.mk
ORCID ID: 0000-0002-0587-1222

Strengthening Resilience Against Deepfakes as Disinformation Threats

Abstract
This chapter looks into the loose definition of deepfakes and, in response to the inherently negative context, the authors provide evidence of the positive uses and benefits of the technology used to create deepfakes. In addition, and for balance, the authors also highlight the inherent threats concerning deepfakes along with the technology's possible employment in criminal activity. To get a better understanding of deepfakes, this chapter also looks at the websites and apps dedicated to deepfake creation and identifies the currently available state-of-the-art, open-source tools. Furthermore, it includes information about the creation of a deepfake video by actually creating one. The main aim and contribution of this paper is to strengthen resilience against deepfakes by highlighting the different factors, the associated regulations and legislation in the EU, and the regulatory situation in North Macedonia. At its conclusion, the chapter provides recommendations on how the general public can identify a deepfake video.

Keywords: Deepfake, Disinformation, Threats, Resilience

Introduction

Deepfakes, or AI-generated synthetic media capable of seamlessly altering or creating content, pose a formidable challenge to the authenticity of information and the integrity of public discourse. As these technological marvels evolve, so do the threats they pose to society. It is estimated that 500,000 video and voice deepfakes will be shared on social media sites globally in 2023 alone (Ulmer, Tong, 2023).

There are myriad possible forms of disinformation based on deepfake technologies. Firstly, deepfakes can take the form of convincing misinformation. Fiction may become indistinguishable from fact to an ordinary citizen when confronted with a deepfake video or voice. Secondly, disinformation may be complemented with deepfake materials to increase its misleading potential. Thirdly, deepfakes can be used in combination with political micro-targeting techniques. Such targeted deepfake work can be highly impactful, especially as regards so-called "micro-targeting", an advertising method that allows deepfake producers to send customised deepfakes that strongly resonate with a specific audience. Looking into recent developments in politics and media, the problem of disinformation reveals a very complex challenge. Deepfakes can be considered in the wider context of digital disinformation, alternative facts, and changes in journalism (Van Huijstee et al., 2021). Deepfakes may also exacerbate social divisions, civil unrest, panic and conflicts, and undermine public safety and national security. In the worst case scenario, this could cause violent conflicts, attacks on politicians, governance breakdown, or threats to international relations (Chesney, Citron, 2018).

As we all confront the challenges posed by deepfakes, it becomes paramount to forge a collective understanding and commitment to fortify our defenses. By fostering resilience and proactive measures, we aspire to safeguard the foundations of truth, trust, and informed decision-making in an age where reality is increasingly shaped by the algorithms of synthetic media.

In this chapter, the authors will look at the definitions of what deepfakes are and also at the negative context usually associated with them, but point out that the technology used for creating deepfakes can serve positive purposes. The available websites and apps for creating deepfakes will be looked at as well as the open source tools and their advantages. To understand the creation process, the authors will create a deepfake and provide their understanding of how ordinary members of the public can learn to identify a deepfake video when they see one.

Definition, Context, and Usage

On a technological level, deepfakes use deep learning as part of AI and enable face swapping with a combination of facial expressions. According to the Merriam Webster dictionary, a deepfake is "an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said" (Merriam-Webster, n.d.). Since the term "deepfake" is loosely defined, there is research on the holistic multidisciplinary definition of deepfakes (Whittaker et al., 2023), a comprehensive overview of deepfakes covering multiple important aspects of different definitions (Altuncu, Virginia, Li, 2022) as well as the need for a more concrete definition (Cochran, Napshin, 2021).

In a research paper entitled, "Tackling Deepfakes in European Policy" (Van Huijstee, et al., 2021) deepfakes are defined as manipulated or synthetic audio or visual media that seem authentic, and which feature people who appear to say or do something they have never said or done, and which are produced using artificial intelligence techniques, including machine learning and deep learning. Deepfakes are most accurately perceived as a subset within the broader classification of AI-generated "synthetic media", including video and audio, photos, and text.

Benefits and Positive Uses of Deepfake Technology

Just by looking at the definition, one can easily garner the negative context around the term, but the technology behind it also has positive potential. There are many available tools that can be used by the public where face swapping is performed with humorous intent, and when friends share the results of their deepfake creations with each other, or where people can swap movie actors' faces for other famous faces, etc. In a more serious manner, the technology can be used in the news anchoring process by using digital twin avatars that would be able to present the news 24/7. It can also be used in movie production in the reduction of the number of retakes, to age or (more usually) de-age actors, and also to break language barriers and allow for more realistic local content. Within the gaming industry, instead of voice actors, game development studios can combine deepfake tech with text-to-speech technologies to achieve multiple outcomes in a single game. In the advertising realm, it can reduce marketing expenses. This technology can also have multiple uses in the education process and address the need for more modern education. For example, historical figures can be used in order to give a better picture of their actions and speeches.

Those who have utilised contemporary smartphones for photography likely have encountered advantages stemming from fundamental deepfake technologies. Frequently, camera applications come with so-called "beauty filters" that automatically alter images to make the subjects look more attractive. More sophisticated deepfakes, involving complete face swaps or speech modifications, can also be created legally, serving purposes such as delivering critical commentary, creating satire and parodies, or simply entertaining an audience. There are evident opportunities for constructive applications of deepfakes in areas such as audio-visual productions, interactions between humans and machines, video conferencing, satire, personal artistic expression, and medical treatment or research.

Deepfake Threats and Criminal Activities

Within the negative context of the aforementioned definition, there are multiple threats that can be initiated, amplified, or combined with deepfakes.

Some of the threats of deepfakes are:

- Deepfakes being used for disinformation;
- The potential for individual defamation through the creation of videos of a victim saying things he/she has never said;
- Identity theft;
- Deepfakes being used for scams whereby the faces of celebrities or popular personas are used to promote products or services;
- AI generated or manipulated content that can affect or change political discourse.

Only a few days after the Russian invasion of Ukraine, a deepfake video of President Zelensky appeared wherein the president appeared to announce his surrender and asks the Ukraine forces to lay down their weapons (Simonite, 2022). In this case, it was obvious that Ukraine had both foreseen and prepared a strategy against this type of attack, and official channels rubbished the deepfake video within minutes of its release. There was also a deepfake video of President Putin in which he declared martial law and called for general mobilisation. This video was broadcast on several Russian radio and television networks (Sonne, 2023).

Deepfake videos can pose a significant threat when combined with other forms of criminal acts. The case of Indian investigative journalist Rana Ayyub serves as a good example, in which an attack on her first started with the creation of fake social media profiles. A deepfake was then created where her face was depicted in a pornographic video. That video was initially shared on social messaging apps such as WhatsApp,

but the largest magnitude of viral activity occurred when a Facebook fan page of India's Bharatiya Janata political party shared the video which resulted in over 40,000 additional shares. The last vector of attack came in the form of Ayyub being doxed, i.e., both her phone number and address were made publicly available (Ayyub, 2018).

Bearing in mind the last example, we come to a situation where the dangers of this type of content are really emphasised. In various countries, video footage may be considered a form of evidence, but the authority and integrity are usually greater when the videos come from video surveillance systems. Like any system, a surveillance system can be a target for a cyberattack, so the danger of deepfakes being inserted into surveillance systems and portraying an innocent person committing a crime can be one of the biggest threats to individuals.

Cyber-based violence represents another form of abuse of women and girls, which is embedded in the gendered social structure and power relations. "The violent acts taking place through technology are an integral part of the same violence that women and girls experience in the physical world, for reasons related to their gender" (GREVIO, 2021).

Technology-facilitated abuse is used as a tool to silence individuals, and also to limit the freedom of speech and human rights advocacy. In most cases, women who are in public and political roles are targeted by campaigns of disinformation, with an intent to discredit, humiliate, intimidate, and silence them in public life (DCAF, 2021, p. 9). Women who are high public figures are often victimised online (Al-Nasrawi, 2021). Powell and Henry (2017) frame sexual violence in cyberspace as "technology-facilitated sexual violence" and define it as an act where information and communication technology are used "to facilitate or extend sexual and gender-based harm to victims" (Powell, Henry, 2017, p. 205). Such terms and definitions give a broader understanding of gender-based violence in the digital space. It is a concept that refers to criminal, civil, or any other type of harmful sexually aggressive, and harassing behaviour being committed with aid or use of technology (Powell, Henry, 2017). Sadly, most of the deepfake content uploaded on the Internet is used for non-consensual pornography, with 98% of all deepfake videos online being pornographic content, of which 99% are women (www.homesecurityheroes.com, n.d.).

Deepfake Software

There are multiple types of software that can be used to create a deepfake, such as DeepSwap, Facemagick, SwapStram, Reface, FaceApp, and Faceswapper among others. Some of these are available as websites,

whereas some are available as iOS and/or Android apps. These websites and apps are mostly used for fun, entertainment, or satirical purposes and charge end users a fee in the form of credits or tokens for more options, datasets, and advanced AI manipulation. Although the end results are to the expected level for their purpose, more realistic and convincing deepfakes are created with open-source tools. Open-source software allows for anyone to view the code, understand how the tools work, and discover any vulnerabilities. Advanced users can edit the code, make modifications, and bug fix. There is also the cost aspect; apps and websites usually charge the end users, whereas open-source tools are free of charge. These aspects are complemented by the community of the open source projects helping other users. The two most popular software used to create deepfakes are Faceswap and DeepFaceLab. Both are Python-based and use deep learning frameworks. They are open source and available with a GPL 3.0 license. GitHub stats such as the number of "stars" (project attributes), the number of people watching, as well and the number of "forks" (new repositories which share code and visibility settings with the original upstream repository) prove these are the most popular deepfaking tools available at the moment. Although there are projects such as DeepFaceLive, from the same developer as DeepFaceLab, Facefusion, SimSwap, and others, Faceswap and DeepFaceLab are far more powerful and have larger communities. These tools come with training processes on multiple images that, most of the time, are extracted from a source and a target video.

Creating a Deepfake

To develop a deeper knowledge of how deepfake videos are made, the authors looked for a tool with set criteria to use, i.e., an open-source tool that can be used with as little expertise as possible. Faceswap and DeepFaceLab, although powerful, have a steep learning curve, so the authors chose another open-source tool called "roop". They used a video where a Prof. Ilijevski gave an interview for the Voice of America (BOA, 2020). From the 720p video, with a total length of 3:31 min, a 33-second portion was clipped. This video was set as a target, and for a source, the authors cropped the head from a photo of Prof. Nenovski. The source image had a resolution of 215 x 241 pixels.

Deepfake creation can be local, i.e., on a creator's PC, or created in the cloud. In the authors' case, they created their video in the cloud and the entire processing took just under 14 minutes. Below we can see a screenshot of the final video.

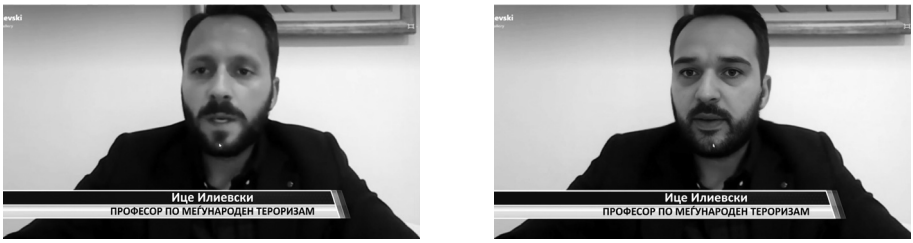

Image 1. Screenshot from the original video (left); and the deepfaked version (right)

From the obtained result, one can see that the final result is a fairly realistic video. In a direct comparison, the authors noticed a greater number of face details in the deepfake video compared to the original video. Here, we have to keep in mind that the authors had access to both the original and the deepfake video for comparison, and in most cases, the public would view the manipulated video without reference to the original video. In the process of manipulating their video, the authors changed the facial features from the forehead to the chin. They did not alter the audio, although that is possible with the aid of additional AI voice manipulation or a provided target video.

If, as demonstrated, and within a brief timeframe and with restricted resources, the authors can achieve significant outcomes using only a single image as the source, it raises well-founded concerns about the influence of powerful disinformation centres. These hubs possess substantial resources, including hardware, software, expertise, and human resources, which amplifies the potential for widespread disinformation and creates a basis for apprehension.

Strengthening Resilience Against Deepfakes

Strengthening resilience against the pervasive threat of deepfakes demands a comprehensive strategy that spans technological, legislative, and societal dimensions. The authors believe there are four pillars for strengthening resilience against deepfakes. These are: raising public awareness; building, implementing, and using better recognition tools; the media and social media company policies; and the government's regulation and legal framework. These pillars are all interconnected because higher levels of awareness can lead to better recognition tools and vice versa. Recognition tools can be created by different entities but the tools created by media and social media companies can have the best vertical integration with their products. Those companies can be

stimulated or pressed with better government acts, bills, and laws which would again lead to better tools and raise public awareness. Better public awareness can be achieved with educational campaigns, more media coverage, workshops, and public service announcements. Bearing in mind the difference in demographic, social, and cultural factors, this process has to be implemented with various media channels, social platforms, and techniques to reach a wide spectrum of target audiences.

Robust legal frameworks are imperative; ones which outline clear responsibilities and consequences for those involved in the malicious creation or dissemination of deepfakes. Collaboration at the international and industry levels is essential, fostering information sharing, research, and the development of innovative countermeasures. Continuous research and innovation, along with user empowerment through controls and transparency, round out the multifaceted approach required to fortify society against the insidious influence of deepfakes.

Regulation and Legislation in the European Union

Since deepfakes can be used as a vehicle for disinformation, the legal framework related to disinformation is also relevant in this context. The creation of a deepfake typically involves the use of personal data, and a deepfake that depicts a natural person can be considered personal data since it relates to an identified or identifiable natural person. Personal data may only be processed under certain conditions since every individual has the right to privacy and data protection. The general rules for processing personal data are laid down in the General Data Protection Regulation (GDPR) (Intersoft Consulting, 2013). The GDPR provides that the processing of personal data always requires a legal basis, and also provides significant directives for addressing illicit deepfake content and grants individuals the right to rectify inaccurate information or have it removed. In each Member State, there exists at least one autonomous supervisory authority tasked with ensuring and enforcing compliance with the established rules and regulations.

In 2018, the European Commission introduced the principles of its strategy to counter disinformation. This strategy encompassed a range of coordinated initiatives across various domains, including enhancing media literacy, bolstering support for high-quality journalism, improving transparency and accountability in online platforms, and safeguarding the online privacy and personal data of citizens. One of the key instruments of the European approach to tackling disinformation online is the Code of Practice on Disinformation (European Commission, 2022a). The Code

was initially set up as a form of self-regulation for the leading online platforms, advertisers, and advertising industry that have committed to: 1) improving the scrutiny of advertisement placements to reduce revenues of the purveyors of disinformation; 2) ensuring transparency with regard to political and issue-based advertising by identifying sponsors and amounts spent; 3) marking automated accounts (bots); 4) empowering users through the promotion of media literacy and providing greater visibility of trustworthy content; and 5) enabling the academic research community to access platform data so that it can track disinformation online and understand its impact.

The 2022 Strengthened Code of Practice on Disinformation (European Commission, 2022a) builds on the pioneering 2018 Code while setting more ambitious commitments and measures aimed at countering online disinformation. The latest Code assembles a broader array of participants than before, enabling them to play a role in comprehensive enhancements by committing to specific obligations pertinent to their respective domains. These commitments encompass measures such as preventing the spread of disinformation, ensuring transparency in political advertising, fostering collaboration with fact-checkers, and facilitating researchers' access to data.

The Digital Services Act (DSA) (European Commission, 2022b), a landmark regulation for the protection of rights in the digital environment, entered into force on 16th November, 2022, and will be directly applicable across the EU from mid-February of 2024. As regards the obligations for very large online platforms and very large online search engines, the DSA starts applying even earlier. The Act contains a set of rules requiring tech companies to properly assess and mitigate the harm their products may cause, as well as to make such assessments and harm mitigation measures available for scrutiny by independent auditors and researchers. As the DSA pertains to content on social media platforms, its relevance extends to the distribution of deepfakes.

Near-simultaneously with the unveiling of the Digital Services Act proposal, the European Commission introduced the European Democracy Action Plan (EDAP) (European Commission, 2021) in December 2020. This Action Plan aims to enhance the resilience of democratic societies within the EU by: 1) promoting free and fair elections; 2) strengthening media freedom; and 3) countering disinformation. At the core of the European approach to tackling disinformation is cooperation between different actors at national and European levels, as well as a multidisciplinarity of responses. This is why the European Digital Media Observatory (EDMO) (digital-strategy.ec.europa.eu, n.d.) was established in June 2020.

Special attention is now being paid to a situation concerning a new European policy for digital strategic autonomy. Strategic autonomy as an imperative requirement would force the EU to expedite its development of critical digital technologies. Other than the need to secure data protection and intellectual property, there is also the need to secure a defense against disinformation (Benedicto-Solsona, Czubala-Ostapiuk, 2023). Indeed, the European Parliament has actively participated in endeavours across the EU to safeguard democratic elections from manipulative interventions and disinformation. Moreover, it has implemented specific measures to address the adverse impacts of artificial intelligence through the adoption of various resolutions and reports.

The latest and most comprehensive document with regard to the discussion of the deepfakes issue is the resolution of 19th May, 2021, on "Artificial Intelligence in Education, Culture and the Audiovisual Sector" (www.europarl.europa.eu, n.d.). This resolution puts forth several proactive suggestions. These encompass the significance of heightening awareness about the risks associated with deepfakes and enhancing digital literacy. It also addresses the growing challenge of identifying and labeling false or manipulated content through technological methods. The resolution urges the Commission to establish suitable legal frameworks governing the malicious creation, production, or distribution of deepfakes. Additionally, it advocates for the advancement of detection capabilities and an enhancing of transparency on the content displayed to platform users, providing them with increased autonomy to decide upon the information they wish to receive.

Countering Disinformation in North Macedonia

In North Macedonia and the Western Balkan region, disinformation campaigns driven by foreign malign influence fluctuate in their frequency, aligning with the prevailing political conditions in the region or a specific country within it. Although the intensity and nature of these campaigns have varied over recent years, addressing diverse potentially divisive issues at any given moment, there has not been a period of complete cessation.

Disinformation represents a significant challenge for North Macedonia, impacting the country's political and social dynamics, as well as public health and safety. Acknowledging this threat, the current government has prioritised the fight against disinformation. In 2019, the Prime Minister publicly introduced the Government's "Plan for Resolute Action against the Spreading of Disinformation", consisting of various, non-binding activities aimed at combating disinformation.

As disinformation campaigns gain momentum, particularly in the context of the COVID-19 pandemic, there is a pressing need to update the "Plan for Resolute Action against the Spreading of Disinformation" to address emerging challenges. To ensure the plan remains relevant, the Government should engage in open consultations with pertinent stakeholders, including media organisations and civil society. The Government has, however, taken the lead in addressing disinformation and hybrid threats more broadly. In October 2021, it adopted the "Strategy for Building Resilience and Tackling Hybrid Threats", accompanied by a 2021–2025 Action Plan. This Action Plan incorporates parliamentary oversight activities and recommends communication channels between informal parliamentary groups and civil society.

A Metamorphosis survey from 2022 of a nationally representative sample shows that over 83% of the respondents agreed with the statement "Disinformation is very harmful and has to be sanctioned by law" (50.8% strongly agree and 32.3% mostly agree). Moreover, 90.8% of the respondents said that "the Government needs to take measures to deal with disinformation in the media" (MetaMorphosis Report, 2022). In the same research, "Citizens identify politicians (91% of respondents), journalists/media (90%), social media (81%), and internet portals (78%) as the main sources of disinformation. In their opinion, the three most important measures to deal with disinformation include: 1) journalists adhering to their professional standards and minding the truthfulness of the content they publish (79%); 2) adopting a law against disinformation in the media (74%); and 3) continuous reporting about the harmful influence of disinformation and fake news in the media (62%)".

How Can We Recognise Deepfake Videos?

In order to familiarise themselves with the convincing level of realism for this type of content, the authors searched for and watched hundreds of examples of deepfake videos. As a result, they went through and looked at such examples available on YouTube and Vimeo, as well as videos embedded in web pages. The available videos are usually not part of academic nor professional research and most of the time only the video is available, and is without any information on the production's used software and tools, the available resources, the amount of data that was used as a source, as well as the time spent training the models. However, viewing numerous videos with different levels of realism allowed the authors to get a clear idea of the state of deepfake videos. It should be noted that in their intentional search for this specific type of content, the

authors were ready for its manipulations. Such an approach cannot be expected from the general public.

From their experience, defined by the subjective factor, they can offer the following recommendation:

- **Intuition:** intuition can be a sign of a critical approach to this type of content. If there are elements in the video that question the validity of the video, they can be a sign that the video is a deepfake.

And the following are the specifics and details that can point us to a video that has been manipulated with AI:

- **Light source:** by identifying the light source, one can review the consistency and the placement of the shadows on a face in shot in relation to the shadows in the background. State-of-the-art software already offers convincing results, but there are instances where the shadows of the face do not correspond with the shadows available in the neck area.
- **Blurred or pixelated parts:** one of the anomalies can be the blurred or pixelised parts of the face. These parts are mostly positioned around the cheek areas where there is less detail compared to more detailed elements of the face such as the eyes, the eyebrows, the nose, or the mouth. We should state that this anomaly is not permanent but can appear temporarily in a video.
- **Facial details compared to background:** deepfake software collects data, builds a model, and inserts another person's face, but the background is not subject to manipulation. The end results may have less detail on the face compared to the background, but there is also some software that inserts another layer of an enhancing process, so in such cases, there can be significantly more detail on the face compared to the background. In such cases, the difference between face versus background details is different from the depth element that is obtained from the cameras themselves.
- **Face details with multiple persons:** if there is a difference in the level of facial detail on different persons in different successive scenes in videos that include multiple persons, such as interviews, it can be symptomatic of a deepfake.
- **Eye blinking:** one indicator for recognising a deepfake video is the intensity of eye blinking. In certain situations, there may be a prolonged lack of eye blinking, and in others, there may be frequent, unnatural blinking.
- **Eye movement:** natural eye movement should be in coordination with facial expressions, body posture, and the message being sent by the speaker. In an AI-generated video, this coordination may

not be retained, especially in situations where the head is turned at a greater angle and the position of the eyes remains towards the person being addressed.
- **Pupils:** another anomaly in AI-generated content is irregular pupil shapes. This is much easier to detect in pictures, but it is not as easy with regard to videos.
- **Reflections in the eyes:** the eyes are the most reflective part of the face. Within different environments, reflections in the eyes can be an indicator of a deepfake. As in the case of the pupils, this is much easier to spot in a picture compared to that of a video.
- **Audio quality:** a video with high-quality visuals, but low-quality audio, may indicate a manipulated video.
- **Background sounds:** additional sounds in addition to the sound from the speaker can be compared to visual elements occurring in the video and one should check whether the background sounds – or lack thereof – are natural to the speaker's environment.
- **Mouth movement:** mouth movement is currently the largest indicator of deepfake videos. Motion can be unnatural for the content reproduced in audio form. Also, certain mouth expressions when speaking, such as the type and intensity of a smile, can betray a manipulated video.

Conclusions

Groundbreaking advancements in AI, particularly Generative Adversarial Networks (GANs), have given rise to deepfakes; altered or synthetic audio and visual content that appears genuine. Presently, smartphone applications with user-friendly interfaces empower individuals to create relatively convincing deepfakes without the need for technical expertise. While the creation of high-quality deepfakes that are virtually undetectable to the human eye, i.e., nearly identical to the real thing, currently demands considerable technical proficiency and specialised equipment, it is anticipated that this requirement may evolve, or, rather, devolve in the foreseeable future.

In this paper, the authors have identified numerous malicious – as well as beneficial – applications of deepfake technologies. The use of deepfake technologies becomes problematic when a creator intends to deceive an audience with malicious intent or influence. The authors conclude that the risks posed by deepfake technologies to society are significant, yet contingent on specific contexts. Given their dual-use nature, these technologies should be subject to regulation.

The US 2024 elections will surely mark an important moment in strengthening not just the USA's resilience against deepfake threats; they can and probably will have global implications. Deepfakes can potentially manipulate public opinion and compromise electoral integrity, so the American elections of 2024 could turn out to be a good moment for legislative efforts and innovative solutions from companies to emphasise the urgency of countering deepfakes. The decisions that will be made during this electoral period will not only shape the USA's resilience but also set a precedent for global approaches in addressing the broader impact of evolving technological threats.

Microsoft, in an anti-deepfake initiative in order to prevent the spreading of disinformation in the US's 2024 elections, has introduced content credentials as a service tool (Hutson, Smith, 2023). Their approach is to use digital watermarking to provide information about the origin of images and videos and determine whether AI has been anywhere near them. In this initiative, Microsoft offers both cybersecurity advice and support to political campaigns. The legal perspective of this initiative is mirrored by the company expressing support for the Protect Elections from Deceptive AI Act and by advocating for legal changes. In the case of Meta, after banning political campaigns from using their generative AI advertising products (Paul, 2023), they also implemented a policy by which they would require disclosure of AI-generated or altered content in political and electoral ads (Kelly, 2023).

In the realm of deepfakes, pursuing legal action as one of its victims can be particularly difficult. Frequently, identifying the perpetrator of one's attack is a serious challenge, as attackers often operate under the veil of anonymity. Additionally, victims may find themselves without the necessary resources to initiate legal proceedings, rendering them susceptible and exposed.

Deepfake technology is a rapidly evolving field, making it challenging to accurately anticipate its future trajectory. Nevertheless, it is certain that visual manipulation is a persistent presence. Quick solutions are currently unavailable, and effectively addressing the risks associated with deepfakes necessitates ongoing contemplation and perpetual learning.

References

Altuncu, E., Virginia and Li, S. (2022) *Deepfake: Definitions, Performance Metrics and Standards, Datasets and Benchmarks, and a Meta-Review*. arXiv, Cornell University. DOI: https://doi.org/10.48550/arxiv.2208.10913.

Ayyub, R. (2018) *I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me*. HuffPost UK. Available at: https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316 (Access 9.10.2023).

Benedicto-Solsona, M.A. and Czubala-Ostapiuk, M.R. (2023) "Rethinking strategic autonomy in times of next-generation EU: New digital agenda", *Journal of Liberty and International Affairs*. Vol. 9(1), pp. 35–47. DOI: https://doi.org/10.47305/jlia2391035s.

BOA (2020) *Ilijevski: makedonskite treba da im pomognat na avstriskite vlasti vo istragata za napadot vo Viena*. Available at: https://mk.voanews.com/a/5651052.html (Access 9.10.2023).

Chesney, R. and Citron, D. (2018) *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. SSRN Scholarly Paper Rochester. NY: Social Science Research Network.

Cochran, J.D. and Napshin, S.A. (2021) "Deepfakes: Awareness, Concerns, and Platform Accountability", *Cyberpsychology, Behavior and Social Networking*. Vol. 24(3), pp. 164–172. DOI: https://doi.org/10.1089/cyber.2020.0100.

digital-strategy.ec.europa.eu (n.d.) European Digital Media Observatory. Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/policies/european-digital-media-observatory (Access 9.10.2023).

European Commission (2021) European Democracy Action Plan. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan_en#documents (Access 9.10.2023).

European Commission (2022a) *2018 Code of Practice on Disinformation. Shaping Europe's digital future*. Available at: https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation (Access 9.10.2023).

European Commission (2022b) *The Digital Services Act package. Shaping Europe's digital future*. Available at: https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package (Access 9.10.2023).

GovTrack.us. (n.d.) *Protect Elections from Deceptive AI Act (S. 2770)*. Available at: https://www.govtrack.us/congress/bills/118/s2770 (Access 9.10.2023).

Hutson, T. and Smith, B. (2023) M*icrosoft announces new steps to help protect elections*. MetaMorphosis Report (2022) *The effect of disinformation and foreign influences on the democratic processes in North Macedonia*. MetaMorphosis foundation Research Report. Available at: https://metamorphosis.org.mk/wp-content/uploads/2022/05/eng-v-7-1.pdf.

Intersoft Consulting (2013) *General Data Protection Regulation (GDPR)*. Available at: https://gdpr-info.eu (Access 9.10.2023).

Kelly, M. (2023) *Meta to require political advertisers disclose AI-generated content*. The Verge. Available at: https://www.theverge.com/2023/11/8/23951346/meta-political-ads-ai-artificial-intelligence-advertising (Access 9.10.2023).

Microsoft On the Issues. Available at: https://blogs.microsoft.com/on-the-issues/2023/11/07/microsoft-elections-2024-ai-voting-mtac/ (Access 9.10.2023).

Merriam-Webster (n.d.) *Definition of DEEPFAKE*. Available at: https://www.merriam-webster.com/dictionary/deepfake (Access 9.10.2023).

Paul, K. (2023) *Exclusive: Meta bars political advertisers from using generative AI ads tools. Reuters*. 7.11.2023. Available at: https://www.reuters.com/technology/meta-bar-political-advertisers-using-generative-ai-ads-tools-2023-11-06/ (Access 9.10.2023).

Powell, A. and Henry, N. (2017) *Sexual Violence in a Digital Age*. Melbourne: Palgrave Macmillan.

Simonite, T. (2022) "A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be", *Wired*. Available at: https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/ (Access 9.10.2023).

Sonne, P. (2023) "Fake Putin Speech Calling for Martial Law Aired in Russia", *The New York Times*. 5.06.2023. Available at: https://www.nytimes.com/2023/06/05/world/europe/putin-deep-fake-speech-hackers.html (Access 9.10.2023).

Ulmer, A. and Tong, A. (2023) "Deepfaking it: America's 2024 election collides with AI boom", *Reuters*. 31.05.2023. Available at: https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/ (Access 9.10.2023).

Van Huijstee, M., Van Boheemen, P. and Das, D. (2021) *Tackling deepfakes in European policy*. European Parliament.

Whittaker, L., Mulcahy, R., Letheren, K., Kietzmann, J. and Russell-Bennett, R. (2023) "Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda", *Technovation*. No. 125, p. 102784. DOI: https://doi.org/10.1016/j.technovation.2023.102784.

www.homesecurityheroes.com. (n.d.) *2023 State Of Deepfakes: Realities, Threats And Impact*. Available at: https://www.homesecurityheroes.com/state-of-deepfakes/ (Access 9.10.2023).

www.europarl.europa.eu (n.d.) *Texts adopted – Artificial intelligence in education, culture and the audiovisual sector*. 19.05.2021. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_EN.html (Access 9.10.2023).