Evolution of Cybercrime and Deepfakes - Exploring Intervention Strategies of International Organizations against AI Threats

By

Priyal Khapra
priyalkhapra@gmail.com

Table of Contents

eISSN1303-5150

Abstract

Deepfakes are basically digital content generated using machine learning (ML), a kind of AI or artificial intelligence, which has generated a lot of interest among common public and marketers and deepfakes are often perceived as "phantom menace" in mainstream media. Irrespectively of being relevant to marketing practice and theory, deepfakes are not fully discussed or understood, along with the opportunities for the deviance or benefits they offer.

Considering the increasing threat of deepfakes on cybersecurity, this paper discusses emerging benefits and threats of deepfakes along with potential solutions. This paper also assesses whether existing international organizations can counter cybercrime to guide the future studies related to deepfakes. It provides a quick analysis of existing policy measures started by international organizations and their relevant impact in the process of making the framework in near future.

It explores the implications of AI and policymaking which would help the existing criminal judicial system to counter cybercrimes. Current uses and AI trends and their applications to do illegal and harmful activities are also analysed. This study concludes by offering various solutions to counter cyber threats committed using AI technologies.

Keywords: *deepfakes, AI, machine learning, data mining, AI trends, international organizations, policymaking*

DOI Number: 10.48047/nq.2022.20.22.NQ10122 **NeuroQuantology 2022; 20(22):1425-1434**

1. Introduction

With the emergence of Artificial Intelligence (AI), deepfake technologies have been developed and posed a huge threat to organizations across the world. Deepfake is an AI technology which can manipulate sounds, images, and video clips to prove an event that never took place (Wojewidka, 2020). For example, a cyber criminal can morph politicians' faces on the bodies of other people who seem to say anything suspicious or offensive that they never said actually. This is a common practice in political situations when cybercriminals have to mislead common public on several debates (Venkataramakrishnan, 2019). For example, an Italian satirical show used deepfake against Matteo Renzi, former PM of Italy.

He was depicted insulting other politicians in the video which went viral on social media. A lot of people started believing on the video as if it was real, which caused outage in public (Venkataramakrishnan, 2019). In addition, deepfakes are also used to impersonate CEOs of companies to cheat account department employees, usually to transfer funds to scammers' bank accounts (Stupp, 2019). Most of the deepfakes are made for entertainment like videogames, movies, and informative videos. Even companies also use deepfakes for marketing and advertisement purposes. For example, companies like Amazon use clips of celebrities to manipulate their words in a way that they are addressing the viewers as per their name, location, etc. to increase sales. However, cybercriminals also make the most of this technology to misguide people and defraud businesses (Westerlund, 2019). In addition, specialized hardware and software and expertise are needed to create those deepfakes. However, even unskilled people can use free tools like "Reface" and "FaceSwap" to manipulate media for scamming individuals and for entertainment purposes (Pfefferkorn, 2019).

1.1 Background

The term "deepfake" combines "deep learning" and "fake" (Rana & Sung, 2020). The phenomenon of deepfake was started on Reddit, a social media platform. A manipulated porn video of a celebrity was shared by an anonymous user and swapped their face with a porn star. Reddit banned the user but their actions increased significant interest in this

phenomenon and this content started spreading on other platforms like 4chan and Twitter (Kirchengast, 2020). Since the onset of deepfakes, they have been used by casual users to manipulate multimedia files by matching human tones and expressions to create media which seems very real to the viewers (Stovold, 2019).

A comedian Jordan Peele depicted Barack Obama giving a speech spreading awareness on the risks of deepfakes, making a common example of using this AI technology (Stovold, 2019). "Generative Adversarial Networks (GANs)" are used to create deepfakes as they use two "Artificial Neural Networks (ANNs)" together. These deep learning models are also called as "synthesizer", "detector", "generative network" or "discriminative network" (Rana & Sung, 2020). They are trained on a vast dataset of images, videos, and sounds to come up with high-end deepfakes (Rana & Sung, 2020).

The synthesizer starts a sequence by creating deepfake which looks authentic enough to trick detector, who is liable to differentiate and analyse whether the clip created by the generator seems to be authentic. The cycle continues until the discriminator cannot detect forgery of the media to improve the quality of deepfake before deploying the same (Westerlund, 2019). GAN algorithms are expected to be more trained on minor datasets and come up with better quality and more convincing deepfakes (Pan et al., 2019). This way, cybercriminals would be able to generate more real deepfakes which would destroy the image of their victims.

2. Literature Reviews

These days, cybercriminals highly misuse deepfakes for illegal activities like extortion, identity theft, spreading fake news, faking obscene videos on celebrities for blackmailing, financial fraud, etc. More than 96% of deepfakes are made of obscene content and they mostly target victims in the US, UK, India, South Korea, and Canada. Cybercriminals created fake audio of a CEO in 2019 to call his company and ask his accounts department to transfer $243,000. Crimes related to deepfakes are increasing every day. Detecting deepfake media is very challenging and is much needed in digital forensics. A cutting-edge approach is needed for the protection of victims against blackmailing with deepfake detection. **Raza et al. (2022)** proposed a novel framework "deepfake predictor (DFP)" to detect deepfake on a hybrid of "convolutional neural network" and VGG16 architecture. The neural network techniques are used to use deepfake dataset on the basis of fake and real faces. They used transfer learning techniques like NAS-Net, Mobile Net, Xception, and VGG16. The DFP framework achieved 94% accuracy and 95% precision to detect deepfakes. Their proposed DFP model performed better than "transfer learning techniques" and other modern studies.

Ahmed et al. (2021)conducted a study to know the awareness of people about Deepfake videos, especially after exposure to those videos. They chose a group of individuals from Bangladesh who was exposed to different deepfake videos. They were asked relevant questions to verify improvement on detection and awareness of deepfake videos. They conducted this study in two stages. The second stage was needed for validation of any generalization. Fake videos are designed for audience and they are made from scratch.

Fake detection has been the need of the hour in this day and age. Deepfakes are powered by "Generative Adversarial Networks (GANs)" which are extended to deep learning to develop amazing capability in audio, speech, and image. It poses a serious threat to individuals and organizations. Deepfake can replace various faces. **Yang et al. (2021)** proposed a smart forensic approach to detect deepfake. They found a subtle texture changes between fake and real images. They used saliency map and guided filter to enhance the artifacts with post-processing and showed the common features of forgery.

Financial crime is very common in cyberspace and several social engineering and hacking attempts have been used by cybercriminals to bypass existing financial information. Financial cybercrime is the new umbrella term which combines hacking, financial crime, and social engineering solely for illegal financial gains. It is not easy to identify financial cybercrimes as a smart algorithm blocks all the suspicious activities. **Nicholls et al. (2021)** conducted a survey to fill the gap by studying the ecosystem of financial cybercrime on the basis of various fraud techniques, relevant algorithms, systems, drawbacks, and metrics to deal with each type of fraud, relevant stakeholders, and emerging and open problems in the domain of financial cybercrime.

Rana & Sung (2020) proposed DeepfakeStack, a deep ensemble learning model to detect deepfake videos. This technique consists of a lot of deep learning-based classification models and makes a smart composite classifier. It performs better than other classifier techniques with over 99.65% accuracy and 1.0 AUROC score to detect deepfake.

2.1 Research Gap

There are plenty of studies conducted to analyse the benefits of AI and machine learning in cybersecurity. There are also several emerging threats which need proper attention from policymakers and organizations. This study is an attempt to take their kind attention towards deepfake, an emerging threat in the field of cybersecurity.

2.2 Research Question

Considering the above arguments, there are some research questions this research paper is aimed to answer –

- What are the implications (threats) of deepfakes for organizations across the world?
- What are the impacts of deepfakes on such organizations?
- How global organizations can mitigate these threats and what are the measures?

2.3 Research Objectives

- To examine the threats of deepfakes on global institutions
- To determine the benefits and applications of deepfakes
- To suggest potential solutions to prevent the threats of deepfakes and mitigate future risks

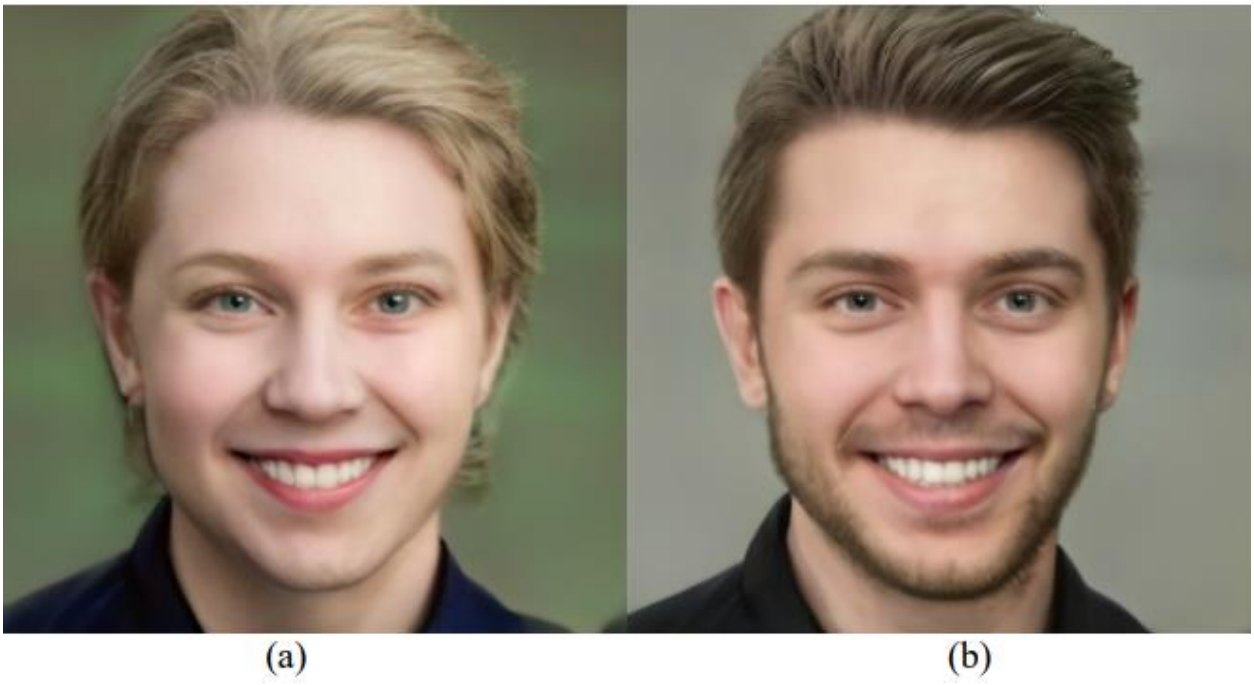
3. Research Methodology

In order to fulfil the above research objectives, this study is based on secondary data gathered from various previous studies conducted on cybersecurity and emerging IT threats based on AI and machine learning. This study has gathered data from relevant studies published in peer-reviewed journals and databases like Google Scholar, Scopus, Research Gate, etc.

4. Analysis of Study

Deepfake is the part of machine learning and AI that merges, replaces, superimposes, or combines different media types to generate a kind of synthetic media which disguises the difference of authenticity (Maras & Alexandrou, 2019). In 2017, Deepfake content first became popular by taking different forms in digital media, be it visual, audio, or audio-visual (Kietzmann et al., 2020). For example, face and body of a person are blended or swapped into someone else's body and face in photographic deepfakes smoothly. In Figure 1, pictures of three people are combined in a way to make deepfake content. An online tool named "Artbreeder" is used to generate these images with deepfake technology, so that users can combine current images and make new ones. The original images of one female co-author and two male co-authors were merged and uploaded (Figure 1), with features of female to be more pronounced in left picture and male features are more evident in right picture (Whittaker et al., 2021).

Figure 1 – Deepfake images of co-authors merged together (a) Female version (b) Male version



Source – Whittaker et al. (2021)

4.1. Threats of Deepfakes on Global Institutions

Deepfake technology poses serious threats to global institutions as they can defraud organizations and increase cybersecurity concerns for them. They can also cause misinformation in the court of law and politics. This section discusses the threats of deepfake technology to judicial systems, politics, and global institutions.

4.1.1. Threats to Legal System

Tampering of evidence is among the serious threats from deepfakes to legal system (Pfefferkorn, 2019). Deepfake can be used to manipulate evidence in the court to get the case in favour of or against the other. Further issues may take place during cross verifications when offering group testifies positively on details of the video while the opposing party is against the video's content (Pfefferkorn, 2019). It would adversely affect the cases as deepfakes might cost money, lead to extra caseloads, and time to authenticate and verify the evidence before being admissible.

For example, a deepfake audio was presented as evidence by a mother in a custody case in the UK (Swerling, 2020). The mother used online tutorials and created deepfake to create a reasonable evidence which felt like father's recording while he was threatening her. She did this to support her claim that the father is abusive and she deserved custody of children. However, with forensic tests, it was proven that audio file was fake and the court dismissed the evidence. Currently, legal systemsare not capable enough to deal with deepfakes used to tamper with the evidences (Maras & Alexandrou, 2019).

According to the "Civil Evidence Act" in the UK, video recordings can be admissible when video is proven trustworthy. However, it is very hard

to differentiate original and deepfake content. In some cases, detection measures couldn't detect deepfakes (Korshunov & Marcel, 2019). Hence, new and effective measures are urgently needed to prevent tampering of evidence in future cases.

4.1.2. Political Threats

Propaganda in politics is another threat that can rise up with deepfakes (Gosse & Burkell, 2020). One can easily create deepfakes and circulate to a huge audience. Deepfake can be used to mislead the public knowingly or unknowingly for spreading political agenda (Gosse & Burkell, 2020). The circulation of a made-up video of Nancy Pelosi, an American politician, on social media was a common example. She looked intoxicated while distorting her words in the video (Reuters, 2019). The former president of the US, Donald J. Trump, posted the clip on his Twitter handle to change public image of Nancy Pelosi, who was his opponent. As a result, the video was shared and viewed on Facebook around 2.5 million times (Greengard, 2019). Irrespective of bilateral calls to take down the video, a spokesperson from Facebook had confirmed that the platform lacks the policies to remove fake information, so the video won't be taken down (Kelly, 2019). Hence, such events had forced governments across the world too look into the matter and regulate deepfakes (Kirchengast, 2020).

In addition, deepfakes can also ruin geopolitical relationships between various countries. Recently, Scott Morrison, the Prime Minister of Australia, called an apology after a spokesperson for foreign ministry of China, Zhao Lijian, for posting a fake image showing a soldier from Australia holding a knife to an Afghan child's throat on Twitter (BBC, 2020). Outrage was sparked on the internet due to that image and bitter debate was started between the Australian and Chinese governments. In addition, this incident might worsen diplomatic relations between both nations. Hence, it is important for politics to control using such deepfakes for political gains on social media.

4.1.3. Corporate Threats

Along with politics and legal systems, deepfakes can adversely affect businesses. With social engineering attacks like phishing, DoS, ransomware, etc., deepfakes can defraud companies which can affect their reputation and internal negotiations (Stovold, 2019). For examples, senior officials can be impersonated by the scammers to achieve fund transfers and sensitive data without getting caught. For example, a UK-based firm was defrauded by the scammers who impersonated their CEO to convince finance department employees for the transfer of $220,000 to their account (Stupp, 2019). A cybersecurity company, Symantec announced that social engineering and deepfakes were used to defraud three "Chief Financial Officers (CFOs)" for undisclosed funding (Sjouwerman, 2019). A financial loss of $250 million was predicted by Forrester Research due to deepfake accounts by the end of 2020 (Sjouwerman, 2019). Businesses might constantly face significant financial losses due to these scams with constant advancement.

Deepfakes can also adversely affect companies using Biometric technology (Wojewidka, 2020). Companies started using biometrics to secure the workplace (Wojewidka, 2020). For example, using face scanners in restricted areas, so that only a few authorized people can enter. In case of breaches due to deepfakes, any unauthorized individual can enter and access intellectual property and sensitive data. This type of attack can cause financial loss because of costs of containing this damage, increasing security, and compensating customers (Rosati et al., 2017).

4.2. Benefits and Applications of Deepfakes

Irrespective of suspicious use, there are also positive applications of deepfakes. For example, machine learning is used by voice assistants like Cortana and Siri. These programs use the same AI models to answer queries of the users and deliver content with voice commands

(Kirchengast, 2020). In addition, "TensorFlow", an AI-based tool, is developed by Google to help discover relevant content in Google Translate and Gmail (Abadi et al., 2016).

Deepfake is also helpful in education sector by providing information in attractive ways to the students (Chesney & Citron, 2019). For example, it can recreate events and figures to provide more immersive experience in subjects like history (Yadav & Salmani, 2019). There are several studies exploring ways to come up with AI program that can automate generating informative content with deepfake. A tool named "LumièreNet" is a classic example to help in making presentations and educational videos on Udacity and other learning platforms (Kim & Ganapathi, 2019).

In addition, deepfake technology improves gaming experience for the players. It can create life-like virtual settings and natural sounding assistants in the game to improve experience (Westerlund, 2019). It is also helpful for filmmakers to cast actors who are not alive like Paul Walker, who died while filming Furious 7. Makers used deepfake technology to recreate his face to shoot the last scene (Yadav & Salmani, 2020). Deepfake has different benefits in social and health care. For example, it can develop a digital copy of loved ones for people to deal with their loss (Westerlund, 2019).

Rehabilitation is another area where deepfake can be used to help people dealing with addiction and substance abuse like smoking. An AI-based program has been developed by the WHO named "Florence" to help people deal with smoking addiction (WHO, 2020). Florence can have virtual conversation with the patients and make them confident to leave their addition by making a plan to track their growth.

4.3. Potential Solutions to Prevent the Threats of Deepfakes and Mitigate Future Risks

A lot of solutions have been deployed and discussed to deal with deepfakes. Currently, deepfakes usually create media contents in low resolution, which one can easily identify with "Convolutional Neural Networks (CNN)" (Li & Lyu, 2018). Li & Lyu (2018) quickly detected and identified deepfakes with 99.1% accuracy. Irrespective of positive results, it is found that one should not rely completely on CNN as there are failure rates in some cases. In addition, existing CNNs would be ineffective soon with large number of excellent-quality deepfakes (Hasan & Salah, 2019).

Digital forensics can be effective to detect deepfakes (Albahar & Almalki, 2019). Forensics can find out whether pixels are altered in the image by keeping anomalies aside like reflections and shadows with computational techniques (Hao, 2018). They can also inspect file metadata to find out any alterations by tracking edit history and number of times of compression of the file. However, using expert tools and dedicated team of forensics can be expensive to detect deepfakes (Lee & Un, 2012). This way, digital forensics has been proposed as a service model by Lee & Un (2012). It uses cloud computing to provide best forensics services at affordable price.

In addition, a solution has been proposed by Hasan & Salah (2019) on the basis of transparency and traceability instead of detection. Those solutions act as clear digital signature on media to ensure authenticity with smart contracts and blockchain. This solution depends upon time-series logs to find out the history of media, track where it was used online to know their origins later on (Hasan & Salah, 2019). One can integrate this solution on a web browser to show the authenticity of content online. However, there are drawbacks of this solution which might negate its benefits. For example, this solution would be error-prone as it might trigger false alarms or identify media content as fake wrongly (Muna, 2020). Smart contracts and blockchain are quite new and they might be hard to implement and costly.

Even with having such technological advancements, there is no guarantee that they

would detect deepfakes with 100% accuracy. More R&D is needed in these technologies as deepfakes would not stop evolving (Lyu & Li, 2018). It is vital to have solutions that can prevent these issues with awareness and employee training. According to Westerlund (2019), one can train employees to find out whether the data is falsified or legitimate. For example, businesses can set up a "2-step authentication" policy to promote employees to verify email and call requests or have another employee to verify transfer of funds (Sjouwerman, 2019). Businesses can further improve their security by controlling accessibility of data to videos and images on social media. It would keep cybercriminals from creating deepfakes and using such data (Sjouwerman, 2019).

Meskys et al. (2020) has discussed another solution by suggesting that governments and tech firms must impose regulations and sanctions on creating deepfakes which can be socially harmful to avoid the spread of fake news and defamation (Kietzmann et al, 2020; Yadlin-Segal & Oppenheim, 2021). However, imposing such restrictions on deepfakes can have adverse outcomes on freedom of expression as these rules would be too forceful to censorship (Hall, 2018). So, it is important to avoid implementing blanket of regulations which could affect freedom of expression.

5. Results

Deepfake technologies have been emerged widely with the emergence of AI and posed a huge threat to organizations. Deepfake is an AI technology which can alter sounds, images, and video to impersonate an event that never happened (Wojewidka, 2020). For example, politicians' faces can be impersonated on the bodies who are supposed to utter controversial words that they never said. This phenomenon is a common practice in political scenarios to spread political agenda and mislead the public (Venkataramakrishnan, 2019). AI systems are widely used in different sectors and criminal justice system and authorities have realized their benefits.

However, law enforcement authorities which are using these technologies for criminal investigation are not completely prepared to face legal and technical consequences of usage of AI for malicious and disruptive purposes. Hence, there is a lack of proper evidence to find out whether law enforcement authorities worldwide are well-trained and equipped to gather evidence globally to perform investigations where AI system can perpetrate or commission illegal activities.

In addition, cooperation and coordination with companies and service providers managing AI systems is important to detect its misconduct and abuse. There are several legal and technical challenges as a lot of AI systems depend on internet connectivity to work where usually traffic and subscriber data should perform investigation. Global service providers will also help locate and identify cybercriminals but they need well-coordinated measures and efforts on the basis of national laws and treaties between private entities and law enforcement agencies. It is more important to have strategic partnerships to deal with cybercrime.

6. Conclusion

In a nutshell, constant development of cybercrime has concluded with deepfakes which extremely magnified the risks of conventional frauds. Deepfakes constantly pose several threats like frauds, tampering of evidence, and spreading propaganda in politics. It is possible to adopt existing technical solutions to avoid attacks related to deepfakes. However, it is not necessary to rely completely on these technologies to deal with deepfakes. It is recommended to invest in training and awareness to identify deepfakes on the onset. Governments should pass the law to criminalize using deepfakes for defamation of people. Proper consequences and punishments should be announced for malicious actors.

References

Wojewidka, J. (2020). The deepfake threat to face biometrics. *Biometric Technology Today*, *2020*(2), 5-7.

Venkataramakrishnan, S. (2019). Can you believe your eyes? How deepfakes are coming for politics. *Financial Times*, *24*.

Stupp, C. (2019). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. *The Wall Street Journal*, *30*(08).

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11).

Pfefferkorn, R. (2019). " Deepfakes" in the Courtroom. *BU Pub. Int. LJ*, *29*, 245.

Rana, M. S., & Sung, A. H. (2020, August). Deepfakestack: A deep ensemble-based learning technique for deepfake detection. In *2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom)* (pp. 70-75). IEEE.

Kirchengast, T. (2020). Deepfakes and image manipulation: criminalisation and control. *Information & Communications Technology Law*, *29*(3), 308-323.

Stovold, T. (2019). What does the rise of deepfakes means for the future of cybersecurity?. Retrieved 30 January 2023, from https://www.kaspersky.com/blog/secure-futures-magazine/deepfakes-2019/28954/.

Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F., & Zheng, Y. (2019). Recent progress on generative adversarial networks (GANs): A survey. *IEEE access*, *7*, 36322-36333.

Raza, A., Munir, K., & Almutairi, M. (2022). A Novel Deep Learning Approach for Deepfake Image Detection. *Applied Sciences*, *12*(19), 9820.

Ahmed, M. F. B., Miah, M. S. U., Bhowmik, A., & Sulaiman, J. B. (2021, July). Awareness to Deepfake: A resistance mechanism to Deepfake. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)* (pp. 1-5). IEEE.

Yang, J., Xiao, S., Li, A., Lan, G., & Wang, H. (2021). Detecting fake images by identifying potential texture difference. *Future Generation Computer Systems*, *125*, 127-135.

Nicholls, J., Kuppa, A., & Le-Khac, N. A. (2021). Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape. *Ieee Access*, *9*, 163965-163986.

Whittaker, L., Letheren, K., & Mulcahy, R. (2021). The rise of deepfakes: A conceptual framework and research agenda for marketing. *Australasian Marketing Journal*, *29*(3), 204-214.

Pfefferkorn, R. (2019). " Deepfakes" in the Courtroom. *BU Pub. Int. LJ*, *29*, 245.

Swerling, G. (2020). Doctored audio evidence used to damn father in custody battle. Retrieved 2 February 2023, from https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/#:~:text=%27Deepfake%27%20audio%20was%20used%20in,is%20being%20submitted%20to%20courts.

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, *23*(3), 255-262.

Korshunov, P., & Marcel, S. (2019, June). Vulnerability assessment and detection of deepfake videos. In *2019 International Conference on Biometrics (ICB)* (pp. 1-6). IEEE.

Gosse, C., & Burkell, J. (2020). Politics and porn: how news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, *37*(5), 497-511.

Reuters (2019). Trump retweets doctored video of Pelosi to portray her as having 'lost it'. Retrieved from https://www.reuters.com/article/us-usa-trump-pelosi-idUSKCN1SU2CB.

Greengard, S. (2019). Will deepfakes do deep damage?. *Communications of the ACM*, *63*(1), 17-19.

Kelly, M. (2019). Distorted Nancy Pelosi videos show platforms aren't ready to fight dirty campaign tricks. *The Verge, last modified May*, *24*.

BBC (2020). Australia demands China apologise for posting 'repugnant' fake image. Retrieved 2 February 2023, from https://www.bbc.com/news/world-australia-55126569.

Stovold, T. (2019). What does the rise of deepfakes means for the future of cybersecurity?. *Kaspersky*. Retrieved 2 February 2023, from https://www.kaspersky.com/blog/secure-futures-magazine/deepfakes-2019/28954/.

Stupp, C. (2019). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

Sjouwerman, S. (2019). The evolution of deepfakes: Fighting the next big threat. TechBeacon. Retrieved from https://techbeacon.com/security/evolution-deepfakes-fighting-next-big-threat.

Wojewidka, J. (2020). The deepfake threat to face biometrics. *Biometric Technology Today*, *2020*(2), 5-7.

Rosati, P., Cummins, M., Deeney, P., Gogolin, F., Van der Werff, L., & Lynn, T. (2017). The effect of data breach announcements beyond the stock price: Empirical evidence on market activity. *International Review of Financial Analysis*, *49*, 146-154.

Kirchengast, T. (2020). Deepfakes and image manipulation: criminalisation and control. *Information & Communications Technology Law*, *29*(3), 308-323.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016, November). Tensorflow: a system for large-scale machine learning. In *Osdi* (Vol. 16, No. 2016, pp. 265-283).

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753.

Yadav, D., & Salmani, S. (2019, May). Deepfake: A survey on facial forgery technique using generative adversarial network. In *2019 International conference on intelligent computing and control systems (ICCS)* (pp. 852-857). IEEE.

Kim, B. H., & Ganapathi, V. (2019). Lumi\erenet: Lecture video synthesis from audio. *arXiv preprint arXiv:1907.02253*.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11).

WHO (2020). Quit tobacco today!. [Online]. Available: https://www.who.int/news-room/spotlight/using-ai-to-quit-tobacco.

Hall, H. K. (2018). Deepfake videos: When seeing isn't believing. *Cath. UJL & Tech*, *27*, 51.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat?. *Business Horizons*, *63*(2), 135-146.

Yadlin-Segal, A., & Oppenheim, Y. (2021). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence*, *27*(1), 36-51.

Meskys, E., Kalpokiene, J., Jurcys, P., & Liaudanskas, A. (2020). Regulating deep fakes: legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, *15*(1), 24-31.

Muna, M. (2020). Technological Arming: Is Deepfake the Next Digital Weapon. *UC Berkley*.

Li, Y., & Lyu, S. (2018). Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*.

Hasan, H. R., & Salah, K. (2019). Combating deepfake videos using blockchain and smart contracts. *Ieee Access*, *7*, 41596-41606.

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

Hao, K. (2018). Deepfake-busting apps can spot even a single pixel out of place. *MIT Technology Review.* Retrieved 2 February 2023, from https://www.technologyreview.com/2018/11/01/139227/deepfake-busting-apps-can-spot-even-a-single-pixel-out-of-place/

Lee, J., & Un, S. (2012, October). Digital forensics as a service: A case study of forensic indexed search. In *2012 International Conference on ICT Convergence (ICTC)* (pp. 499-503). IEEE.