Rising Above Misinformation and Deepfakes

Namosha Veerasamy and Heloise Pieterse
CSIR, Pretoria, South Africa
nveerasamy@csir.co.za
hpieterse@csir.co.za

Abstract: Misinformation can be rapidly spread in cyberspace. It thrives in the social media landscape as well as news platforms. Misinformation can readily gain momentum in the race to influence people or intentionally deceive. With the use of bots, misinformation can be easily shared, especially in environments like Twitter and Facebook. While, some measures are taken to stop the spread of misinformation, threats like Deepfakes are posing a higher challenge. Deepfakes provide a means to generate fake digital content in order to impersonate a person. With the use of audio, images and videos, artificial intelligence is used to depict the speech and actions of people. Deepfakes are typically made of presidents or influential businessmen such as Donald Trump and Mark Zuckerberg. Deep Fakes can be very realistic and convincing as this form of synthetic media is raising concerns about its possible misuse. The effects of Deepfakes are to spread disinformation, confuse users or create influence. This can lead to further effects like political factions, blackmail, harassment and extortion. Deepfakes could lead to a distrust in digital content as many may feel that anything we see is actually just a manipulation. Deepfakes has arisen as a new generation of misinformation through the manipulation of digital media in order to create realistic videos. This paper looks at the governing, communal and technical issues relating to Deepfakes. At the technical level, the use of audio and text analysis used to create Deepfake videos is advancing at a rapid pace which has also made its affordability and accessibility easier. An evaluation of the threats stemming from Deepfakes reveals that there are various mental, monetary and group dynamics involved. In this paper, the various types of threats emanating from Deepfakes is discussed. This paper also looks at five factors in the field of Deepfakes that should be taken into consideration (Technical Source Dissemination Victim Viewers). The paper discussed these five factors in order to help identify measures to help curb the spread of Deepfakes. A combination of these measures can help limit the spread of Deepfakes and support mitigation of the threat. Due to prominence and power that digital media has, it is imperative that this threat not be overlooked. The paper provides a holistic approach to understanding the risk and impact of Deepfakes, as well measures to help mitigate abuse thereof.

Keywords: Deepfakes, Misinformation, Voice Cloning, Text Synthesis, Machine Learning

1. Introduction

The manipulation of digital content is not a new phenomonen. However, improvements of machine learning technologies have given rise to a new generation of manipulated digital content – Deepfakes. The term "Deepfake" is an amalgamation of both "deep learning" and "fake" (Anon., 2021) and describes the production of highly realistic manipulated digital content. While most circulated Deepfakes are parodic in nature (Collins, 2019), a few occurrences have illustrated Deepfakes as an ideal platform for propaganda.

The proliferation of Deepfakes are driven by two factors: technological advances and societal context. Technological advances, such as increased computing power, 5G connectivity, 3D sensors, high-quality algorithms and pre-trained models, drive improvements in the quality of Deepfakes. From a societal perspective, the changing media landscape (a shift from the traditional model of centralised information distribution to user-generated content platforms) and the reliance on visual communication will cause Deepfakes to become a prominent source of information in the near future (Anon., 2021).

While the growing concern associated with Deepfakes remaining the distribution of misleading information, it is important to note that Deepfakes also provides potentially beneficial uses, especially for the entertainment industry. In Gemini Man, a 2019 hollywood blockbuster starring Will Smith, the special effects team reportedly relied on Deepfakes techniques to generate a younger but completely digital clone of Will Smith's character (Bradshaw, 2019). Other positive uses of Deepfakes include voice-synthesis for medical purposes, animaton of art to create virtual museums, as well as interactive educational lessons. However, the risks associated with Deepfakes necessitates continuous investigation of the technology.

The purpose of this paper is to investigate Deepfakes from the perspective of the following five factors: Technical, Source, Dissemination, Victim and Viewers. Such a broad understanding of Deepfakes is required to gather the necessary insight to identify applicable characteristics that can assist mitigation strategies. Mitigating the risks associated with Deepfakes, and manipulated digital content in general, requires continuous attention as Deepfake technology is a fast-moving target without any quick fixes currently available.

The remainder of this paper is structured as follows. Section 2 provides an overview of Deepfakes, focusing on the evolution of Deepfake technology, as well as highlighting notable Deepfake occurrences. In Section 3, the five factors of Deepfakes are discussed while Section 4 presents noteworthy characteristics of Deepfakes. Section 5 discusses mitigation factors that can be used to minise the impact of Deepfakes as a form of misinformation and Section 6 concludes the paper.

2. Background

2.1 Evolution of Deepfakes

Commonly viewed by the general public as unltrarealistic fake videos (Harris, 2019), Deepfakes are better described as a subset of broader category of AI-generated "synthentic media", including not only video and audio, but also photos and text (Anon., 2021). Deepfake technology first appeared on Reddit, a popular social media platform comprising of smaller communities used for content sharing and discussion. In November 2017, a user, known only as u/deepfakes, created such a community and shared the first rendition of the Deepfakes algorithm. Shortly thereafter, the community was renamed r/deepfakes, launching a new era of face-swapping videos (Fikse, 2018). The first collection of Deepfake videos contained pornographic content, in which the faces of the original actresses were replaced by well-known celebrities.

The first generation of Deepfake videos were characterised by: (i) low-quality synthesised faces, (ii) different colour contrast among the synthesised fake mask and the skin of the original face, (iii) visible boundaries of the fake mask, (iv) visible facial elements from the original video, (v) low pose variations, and (vi) strange artifacts among sequential frames. As of today, Deepfake videos have evolved both in quality and quantity due to improvements in technology and publicly available databases. The second generation of Deepfake videos considers different acquisition scenarios (e.g., indoors and outdoors), light conditions (e.g., day and night), distances from the person to the camera, as well as pose variations (Tolosana, et al., 2021).

Further advancements in Deepfake technology are demonstrated by the creation Deepfake videos using a single profile picture. Researchers at Samsung's AI lab created a method to train a model using a single photo and various landmark facial features (e.g., shape of the face, eyes, mouth shape, etc.). While glitches were still clear and obvious, the technique elevates the risks of misinformation, deception and fraud to new levels (Solsman, 2019). Further research by Siarohin et al. improved on the technique by using a motion extractor that learns the extraction of key points along with their local affine transformations. Afterwards, a generator network models occlusions in the target motions and combines the appearance extracted from the source image and the motion derived from the driving video. (Siarohin, et al., 2019). The outcome is the animation of a static photo without requiring any prior information or knowledge of facial landmarks.

Technological improvements have enabled other forms of Deepfakes to evolve. Voice cloning technologies, also known as audio-graphic Deepfakes, speech synthesis or voice conversion/swapping (Neelima & Santiprabha, 2020), can generate sythentic speech that closely resembles a targeted human voice. Based on the concept of Text-to-Speech (TTS) technology, voice cloning relies on one of the following approaches: Concatenative TTS and Parametric TTS. Concatenative TTS uses a collection of audio recordings containing words and sounds to formulate sentences while Parametric TTS relies on statistical models of speech to simplify voice creation (Martin, 2021). As of today, advances in AI and deep learning have increased the quality of Parametric TTS-based voice cloning, leading to the creation of easily accessible and reusable AI-powered tools such as Tacotron, WaveNet, Deep Voice, and Voice Loop (Anon., 2021).

Text synthesis technologies relies on Natural Language Processing (NLP) to generate Deepfake text that imitates the unique writing and speaking style of a target. Key to NLP is a deep learning algorithm called the Transformer that allows for the transformation of input text into new text, by learningword sequences and sentence construction. One of the most advanced language models built on this architecture is Generative Pre-trained Transformer 3 (GPT-3), which was developed by OpenAI – an AI research laboratory. GPT-3 is a general-purpose NLP model consisting of 175 billion parameters that demonstrated notable performance related to translation, question-answering, as well as unscrambling words. However, researchers have found that the high-quality text generating capability of GPT-3 can make it difficult to distinguish synthetic text from the human-written text, leading to the potential misuse of language models (Brown, et al., 2020).

2.2 Notable Deepfake Occurrences

Over the past few years, multiple examples of Deepfake videos have appeared on YouTube, a popular video sharing and social media platform. Most Deepfake videos were created solely for entertainment purposes, however, a few key examples illustrate the potential power of the videos as a tool to spread misinformation.

One of the most prevalent and widely shared Deepfake videos stars Jordan Peele as former US President Barrack Obama. The video was created using FakeApp and involved taking an original video of Barack Obama and pasting Jordan Peele's mouth into it. The video, which took roughly 56 hours to develop, was created to raise awareness regarding the potential misuse of Deepfake videos (Silverman, 2018).

An example of a politically-motivated Deepfake video involved another former US President, Donald Trump, encouraging Belgium to withdraw from the Paris climate agreement. The Deepfake video was published by a Belgian social democratic party with the purpose of starting a public debat regarding the necessity to address climate change. The video was eventually debunked by Lead Stories (Lytvynenko, 2018).

In 2019, a video appeared on Instagram falsely portraying Facebook's Chief Executive Officer, Mark Zuckerberg, crediting a secretive organisation for the success of Facebook (Shead, 2020). The video was created by Canny AI, an Israeli company, and constructed using deepfake technology, an actor's voic and 2017 footage of Mark Zuckerberg (Metz & O'Sullivan, 2019).

The evolution of Deepfake technology was further demonstrated by criminals using AI-based software to impersonate the voice of a chief executive officer (CEO) and demand a fraudulent transfer of $243000. The target, a CEO of a UK-based energy firm was convinced the caller was the CEO of the firm's German parent company and, therefore, approved the request. The criminals relied on the principles of authority and urgency to ensure the transfer took place within an hour (Stupp, 2019).

Finally, in 2021, various news outlets reported several European parliament members were targeted by Deepfake video calls imitating Russian opposition. The calls, which were created by Russian pranksters, Vovan and Lexus, were described as hyper-realistic and impressive (Roth, 2021).

3. Five Factors of Deepfake

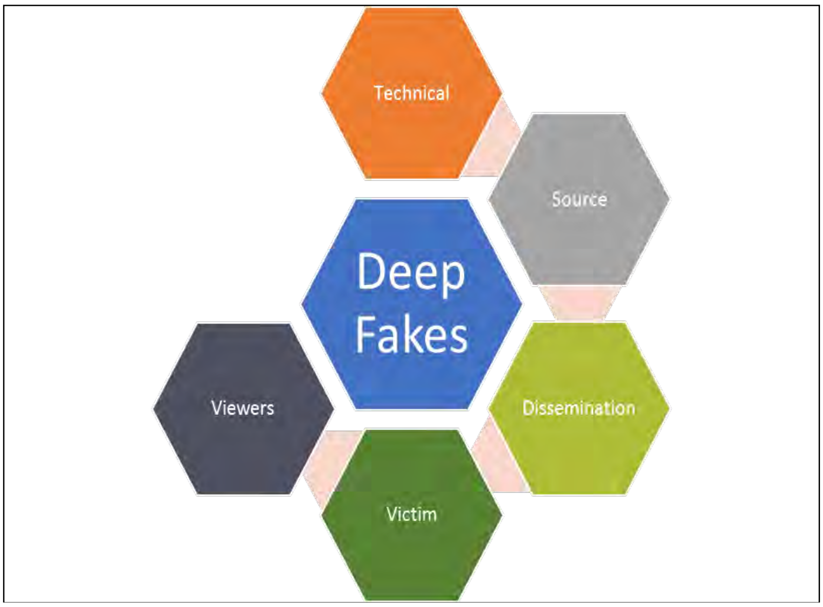

Figure 1: Five factors in field of Deep Fakes

This paper looks at five key factors in the field of Deepfakes that should be taken into consideration-see Figure 1 (Technical Source Dissemination Victim Viewers). The paper discussed these five factors in order to help identify measures to help curb the spread of Deepfakes. A combination of these measures can help limit the spread of Deepfakes and support mitigation of the threat. Due to prominence and power that digital media has, it is imperative that this threat not be overlooked. These five factors are discussed next

3.1 Technical

There are three technological factors contribute that contribute to the current Deepfake phenomenon: (i) AI concepts that leverages powerful machine learning (ML) and deep learning techniques, (ii) increased computing power, and (iii) datasets needed to train the algorithms.

3.1.1 Algorithms

Photo and video-graphic Deepfakes are commonly constructed using Artificial Neural Networks (ANNs), which are computing systems inspired by biological neural networks (Jain, et al., 1996). The basic processing elements of ANNs are a collection of artificial neurons, called nodes, which are connected to form edges and enable transmission between the neurons. Effects on edges are represented by weights (real numbers) and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. Both generative Adversarial Networks (GANs) and Autoencoders are examples of ANNs used to create phto and video-graphic Deepfakes (Abraham, 2005).

GANs were designed by Ian Goodfellow and his colleagues in 2014 (Goodfellow, et al., 2014) and are based on game theory involving two ANNs: generator and discriminator. A 'generator network' analyses a training set and identifies common patters while a 'discriminator network' aims to determine whether the created content is similar by identifying forgeries. Detected forgies are reported and the 'generator network' will attempt to improve the authenticity (Goodfellow, et al., 2020). The methodology provided by GANs have proven to be common use for the development of Deepfakes (Rana & Sung, 2020). Autoencoders represents another form of ANNs that are commonly used to construct Deepfakes. Any Autoencoder consists of two components: encoder and decoder. The encoder extracts the most important features to recreate the original input (e.g., photo of a person). To decode the features, separate decoders are generated for person A and person B, which are trained using a neural network. Once training is completed, face swapping can be performed by replacing the original decoder for person A with the decoder constructed for person B (Hui, 2018).

3.1.2 Computing Power

Successful development of Deepfakes requires appropriate computing power needed to execute the above-mentioned algorithms. For example, the supercomputer developed for OpenAI's GPT-3 is a single system with more than 285,000 CPU cores, 10,000 GPUs and 400 gigabits per second of network connectivity for each GPU server (Langston, 2020). However, the current proliferation of Deepfakes illustrate the increasing availability and affordability of computing resources (e.g., CPUs, GPUs, or high-performance supercomputers). The cost associated with high-performance computing has reduced rapidly, driven by the increasing availability of cloud computing services (Collins, 2019).

3.1.3 Data Requirements

The final technical requirement for generating Deepfakes is an adequate training set. Traditionally, the development of Deepfakes relied on large volumes of data – the more training data available, the better the quality of the Deepfake. The Internet has contributed significantly to the generation of training sets with public figures (e.g., former US Presidents and celebrities) becoming the most favoured targets due to their prominent online digital footprint. However, recent research have demonstrated the possibility of generating Deepfakes using smaller datasets (Siarohin, et al., 2019, Zakharov, et al., 2019).

3.2 Source

The source of Deep Fakes can also be labelled the perpetrator. In many instances, the identity of the source is hidden and thus the source cannot be held accountable for actions taken. Deep Fake footage can be created by technology enthusiasts, researchers, academics, state-sponsored actors, political groups, lone actors, and even pornography developers. In terms of mitigation measures, legal frameworks may need to be developed in order to classify Deep Fakes as criminal acts. Also, diplomatic relations and international agreements may need to include members declaring that such technology will not be used for opportunistic or malicious purposes.

The creators/source of Deep Fakes mainly seek to achieve some of the following objectives:

- Misconstruction of truth
- Spread of fake news
- Mislead
- Creation of shock or alarm

- Discredit high profile individuals
- Entertainment
- Fraud
- Manipulation of events like elections
- Intimidation or blackmail
- Damage to stability or relations

The creators of Deep Fakes mainly seek to create impactful or destructive content that can have a consequential effect on individuals, companies and groups of people. Deep Fakes target individual persons but the effects can felt at a wide-scale level. The cascading resultss can be seen at different levels (see Fig 2).

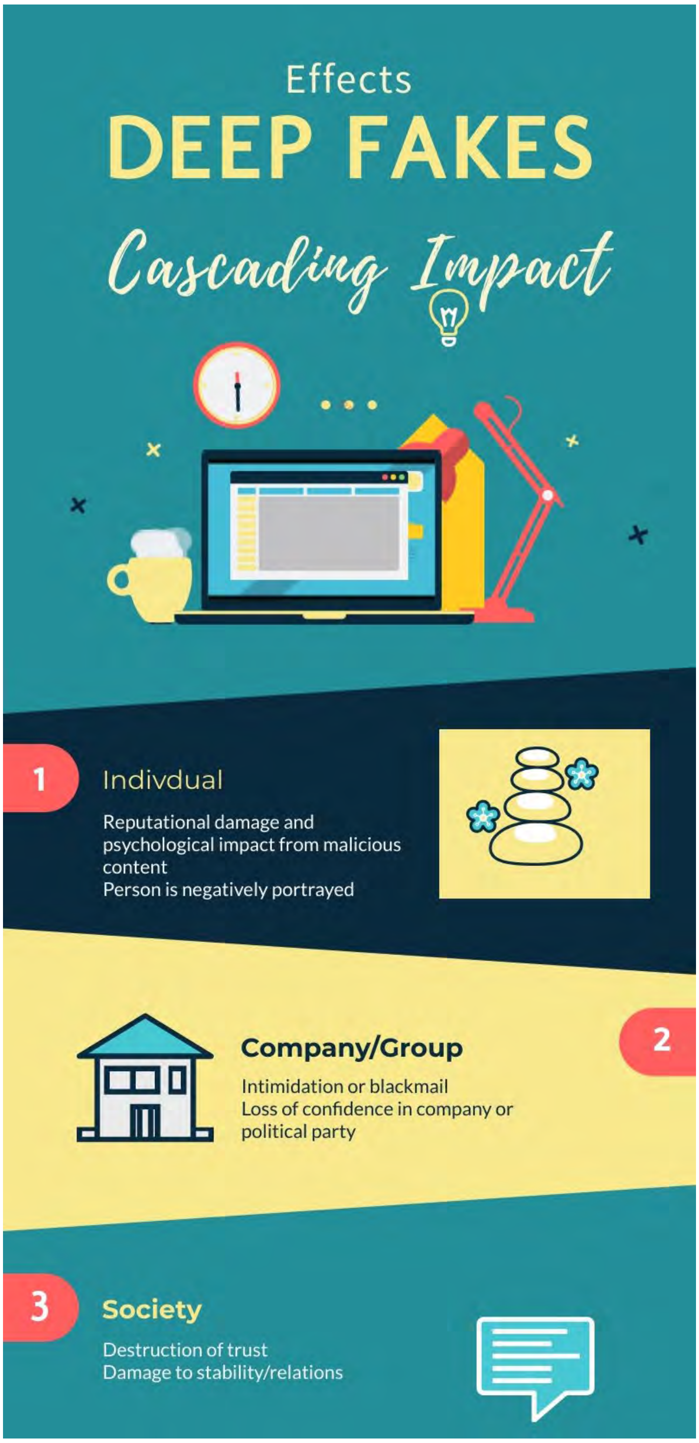
Effects
DEEP FAKES
Cascading Impact
1
Indivdual
Reputational damage and psychological impact from malicious content
Person is negatively portrayed
Company/Group
2
Intimidation or blackmail
Loss of confidence in company or political party
3
Society
Destruction of trust
Damage to stability/relations

Figure 2: Cascading Effects of Deep Fakes

3.3 Dissemination

Social Media (Facebook and Youube) as well as online platforms are the key mechanisms that are used in the dissemination of Deep Fakes. These dissemanation channels play a key role in the scale and severity of impact. It is thus crucial that these online domains take some responsility and carry some obligation in preventing the misuse of Deep Fakes. It is imperative that deepfake detective technologies be deployed, acting on the results as well as taking down identified malicious Deep Fakes. The circulation cycle needs to be adjusted in order to prevent rampant distribution and the negative effects thereafter. One of the most effective solutions may come from major tech platforms like Facebook, Google and Twitter voluntarily taking more rigorous action to limit the spread of harmful Deep Fakes (Toews, 2020). Efficient labelling and take down procedures should be in place.

Facebook enlisted reseachers from Bekely, Oxford and other institutions to build a deepfake detector and help implement a ban as well as Twitter making policy changes to tag any Deep Fakes that are not removed (Adee 2020). In February 2020, Youtube declared that it would also be banning deefake videors relating to the US election, voting process and US census (Adee 2020).

Furthermore two programs called Reality Defender and Deeptrace can be used to block Deep Fakes. Deeptrace uses an APT that will act like a hybrid antivirus/spam filter prescreening incoming media and diverting obvious manipulations to a quarantine zone, similar to how Google diverts spam (Adee 2020). Reality Defencer comes from the company AI Foundation and aims to tag and contain manipulated images and video before they cause damage.

3.4 Victim

A core target of Deep Fakes are women. Many victims of image-based sexual abuse—also known as revenge porn—or deepfakes feel powerless (Fighthenewdrug.org, 2021). Nonconcensual pornography accounts for 96% of deepfakes currently on the Internet (Adee, 2020).

However, women are not the only target. Deep Fakes can be used to bully, imimidate, trick or mock victims.

Many companies may be concerned about Deep Fakes due to the potential use for fraud and theft. Deep Fakes can be used to convince people in power to make payments to fraudsters. Extortion could be another key area of concern. Deepfakes could be used to make fraudulents payments or hack into individuals banking services.

At a national level, Deep Fakes could threaten democracy. Shocking video footage could be created of electoral candidates in order to smear their campaigns. Celebrtites are also targeted and Deep Fakes discredit their reputation. In many caseses, Deep Fakes are created of prominent figures in order create misinformation, manipulate the public, mislead or generate shock or confusion.

From the perspective of protection, establishment of support mechanisms for victims is needed, as well as improving the capacity of data protection authorities together with encouraging the protection of fundamental human rights.

3.5 Viewers

Deep Fakes have a cascading impact from the individual to group dynamics, as well as at societal and community level. Deepfakes threaten to grow from an Internet oddity to a widely destructive political and social force and society needs to act now to prepare itself (Toews, 2020)). Viewers of the Deep Fakes may react in different ways. Some may choose to believe the footage and lose trust in the individual or company. A deepfake can leave an imprint and hurt someone’s reputation especially if their name and face is part ofa negative video or audio-real or deepfake (Somers 2020) In terms of protection, it is important that investment is made in labelling trustworthy sources and media and security awareness so that the public is able to identify potential Deep Fakes.

4. Characteristics of Deepfakes

Carefully made deepfakes can already be very realistic, though only under certain circumstances—an attentive observer will notice that convincing deepfakes focus on individuals who don’t wear glasses or have beards, and typically use a stationary camera (Engler, 2019).

In some instances by studying videos carefully, one may be able to detect that the footage is a Deep Fake. Some characteristics to look for (Kapersky, 2021)

- Jerkey movements
- Shifts in lighting from one frame to the next
- Shifts in skin tone
- Strange blinking or no blinking (or long pauses between blinking)
- Poorly synched lips and cound
- Digital artifacts in the image

Sample mentions some of the following signs to look out for as well (2020):

- Bad lip synching
- Patchy skin tone
- Flickering around edges of transposed faces
- Fine details like hair are difficult to render wll
- Badly rendered jewelery or teeth
- Strange lighting effects
- Inconsistent illumination and reflections on the iris

5. Mitigation Factors

In the security world, there is continuous dynamic of the creation of protection mechanisms which attackers then try to figure out a means to bypass. With Deep Fakes, we find the same paradign. For example, a filter can be added over an image so that the image cannot be used to create a deepfake. However, innovative technologiest may find ways to circumvent controls and still accomplish their goal of Deep Fake generation.

Another protection mechanism is digitally signing content. SImilar to a security certificate, content can be validated through a third party. The challenge here is that in some cases, viewers will not be interested in whether the content is digitally signed. For example, revenge porn.

Another emerging technology is the use of cryptographic algorithms to insert hashes at certain intervals in the video; if the video is modified the hash will change (Kapersky, 2021). This is comparative to watermarking documents.

Another preventative techniques is to use a program to insert specially designed digital artifacts into the the videos to conceal the patterns of the pixels that face detection software uses (Kapersky, 2021). This will slow down Deepfake algorithsm or lead to poor quality videos.

The International Risk and Governance Center has proposed several measures to help govern Deepfake risks. These include (Collins, 2019):

- Detection: Continued research into technologies to distinguish between authentic and fabricated digital content
- Provenance: Techniques designed to verify the origin and integrity of digital artefacts, such as trusted-hardware schemes or ways of preserving metadata
- Image rights and control: Greater control for individuals over digital content that relates to them, including potential “takedown” rights
- Digital corroboration: The use of multiple independent data sources, analogous to the familiar process of corroborating eye-witness testimony
- Secure digital processes: A greater focus on authentication and verification to make digital communication less vulnerable to deepfakes
- Platform nudges: Interventions to influence the way people — and algorithms — share digitalcontent
- Legal guidance- clarification of the ways in which existing legal frameworks apply to deepfakes
- Penalties- The persistence nature of some harms involving digital content may require changes in the way they are penalised

Technology platforms have a key role to play in the spread of Deep Fakes. Efficient labelling and take down procedures should be implemented. A platform can flag a Deep Fake once detected. This can help reduce the spread of harmful content.

At a policy level, measures should be put in place like extension of legal frameworks to include Deep Fakes as criminal offenses, as well as international agreements to refrain from the use of Deep Fakes. Furthermore, more stringent measures like the imposition of sactions on states involved in disinformation and Deep Fakes could help curb the spread.

Furthermore at a research level, tremendous investment is needed to develop AI techniques that can block, slow or obscure Deep Fake creation.

Overall, awareness will also be needed in order to educate the public about the possbiilitieis and dangers of Deep Fakes. An informed community may form a crucial aspect of defending against misinformation from Deep Fakes.

6. Conclusion

It is expected that more Deep Fakes will be distributed in order to manipulate public opinion, carry out distortion, manipulation of elections, undermining journalism,inflicting reputational damage, destabilise and cause mininsformation. The growth of Deep Fakes may erode public faith in digital media. The effects of Deep Fakes will be spread of disinformation, increase in doubt as well as ability to deny accountablibily (claiming material is fake). The public may find it increasingly difficult to differentiate from what is real and what has been maliciously generated. We live in a world undergoing a rise in disinformation and fake news. The modern version of photoshopping , users can now impersonate a politician, appear in a movie or dance like a professional. Deep Fakes may grow rapidly. There are no easy solutions. Overall, more continuous efforts and governance is needed in order to curb the prominent spread of Deep Fakes. Technology providers need to take a more ative role in the proper use of their technologies. Policy makers and regulators need to implement stronger measures in order to protect victims and punish perpetrators. Legal frameworks need to include the criminal implications of malicious Deep Fakes. International agreements should concur that members will refrain from the use of Deep Fakes. Furthermore, an additional consideration is the imposition of sanctions on states engaged in disinformation and Deep Fakes. Visual manipulation has emerging as a fast-moving trend. It is imperative that authorities and specialists around the world put in a concerted effort to mitigate its rampant growth.

References

Abraham, A., 2005. Artificial neural networks. In: P. H. Sydenham & R. Thorn, eds. *Handbook of measuring system design.* s.l.:John Wiley & Sons, pp. 901-908.

Adee S, April 29 2020, What are Deepfakes and how are they created?, IEEE, Available at https://spectrum.ieee.org/what-is-deepfake, [Accessed 27 September 2021].

Anon., 2021. *Tackling deepfakes in European policy,* s.l.: European Parliamentary Research Service.

Brown, T. et al., 2020. *Language models are few-shot learners.* s.l.:arXiv preprint arXiv:2005.14165.

Collins, A., 2019. *Forged Authenticity: Governing Deepfake Risks,* Lausanne: EPFL International Risk Governance Center.

Engler A, November 14 2019, Fighting deepfakes when detection fails, Available at: https://www.brookings.edu/research/fighting-deepfakes-when-detection-fails/, [Accessed 27 September 2021].

Fightthenewdrug.orgviewer, 2021, 7 Things you can do if you're a victim of Deepfakes or revenge porn, Available at : https://fightthenewdrug.org/7-things-you-can-do-if-youre-a-victim-of-deepfakes-or-revenge-porn/, [11 Octboer 2021].

Fikse, T. D., 2018. *Imagining Deceptive Deepfakes.* Oslo: University of Oslo.

Goodfellow, I. et al., 2014. Generative adversarial nets. *Advances in neural information processing systems,* Volume 27, pp. 1-9.

Goodfellow, I. et al., 2020. Generative adversarial networks. *Communications of the ACM,* 63(11), pp. 139-144.

Harris, D., 2019. Deepfakes: False Pornography Is Here and the Law Cannot Protect You. *Duke Law & Technology Review,* Volume 17, pp. 99-127.

Hui, J., 2018. *How deep learning fakes videos (Deepfake) and how to detect it?.* [Online]
Available at: https://jonathan-hui.medium.com/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9
[Accessed 20 September 2021].

Collins, Aengus. (2019). Forged Authenticity: Governing Deepfake Risks. Lausanne: EPFL International Risk
Governance Center.

Jain, A., Mao, J. & Mohiuddin, K., 1996. Artificial neural networks: A tutorial. *Computer,* 29(3), pp. 31-44.

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

Kapersky, 2021, Deepfake and Fake Videos- How to Protect Yourself, Available at: https://www.kaspersky.co.za/resource-center/threats/protect-yourself-from-deep-fake, [Accessed 27 September 2021].

Langston, J., 2020. *Microsoft announces new supercomputer, lays out vision for future AI work.* [Online] Available at: https://blogs.microsoft.com/ai/openai-azure-supercomputer/ [Accessed 20 September 2021].

Lytvynenko, J., 2018. *A Belgian Political Party Is Circulating A Trump Deepfake Video.* [Online] Available at: https://www.buzzfeednews.com/article/janelytvynenko/a-belgian-political-party-just-published-a-deepfake-video [Accessed 18 September 2021].

Martin, K., 2021. *What is Voice Cloning.* [Online] Available at: https://www.idrnd.ai/what-is-voice-cloning/ [Accessed 20 September 2021].

Metz, R. & O'Sullivan, D., 2019. *A deepfake video of Mark Zuckerberg presents a new challenge for Facebook.* [Online] Available at: https://edition.cnn.com/2019/06/11/tech/zuckerberg-deepfake/index.html [Accessed 18 September 2021].

Neelima, M. & Santiprabha, I., 2020. *Mimicry Voice Detection using Convolutional Neural Networks.* s.l., s.n., pp. 314-318.

Rana, M. & Sung, A., 2020. *Deepfakestack: A deep ensemble-based learning technique for deepfake detection.* s.l., IEEE, pp. 70-75.

Roth, A., 2021. *European MPs targeted by deepfake video calls imitating Russian opposition.* [Online] Available at: https://www.theguardian.com/world/2021/apr/22/european-mps-targeted-by-deepfake-video-calls-imitating-russian-opposition [Accessed 18 September 2021].

Shead, S., 2020. *Facebook to ban 'deepfakes'.* [Online] Available at: https://www.bbc.com/news/technology-51018758 [Accessed 18 September 2021].

Siarohin, A. et al., 2019. First Order Motion Model for Image Animation. In: *Advances in Neural Information Processing Systems.* s.l.:Curran Associates, Inc., pp. 1-11.

Silverman, C., 2018. *How To Spot A Deepfake Like The Barack Obama–Jordan Peele Video.* [Online] Available at: https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed [Accessed 17 September 2021].

Solsman, J., 2019. *Samsung deepfake AI could fabricate a video of you from a single profile pic.* [Online] Available at: https://www.cnet.com/tech/computing/samsung-ai-deepfake-can-fabricate-a-video-of-you-from-a-single-photo-mona-lisa-cheapfake-dumbfake/ [Accessed 13 September 2021].

Somers M, July 21, 2020, Deepfakes Explained, MIT Management Sloan School, Available at : https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained, [Accessed 11 October 2021].

Stupp, C., 2019. *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case.* [Online] Available at: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 [Accessed 18 September 2021].

Toews R, May 25 2020, Deepfakes are going to wreak havoc on Society. We are not prepared, Forbes, Available at: https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=2f9027c17494, [Accessed 27 September 2021].

Tolosana, R., Romero-Tapiador, S., Fierrez, J. & Vera-Rodriguez, R., 2021. *Deepfakes evolution: Analysis of facial regions and fake detection performance.* s.l., Springer, pp. 442-456.

Zakharov, E., Shysheya, A. & Burkov, E., 2019. Few-Shot Adversarial Learning of Realistic Neural Talking Head Models. *CoRR.*