International Journal of Computational and Biological Intelligent Systems (IJCBIS)
Volume 03, Issue 02, October 2021

Deepfake Generation, Detection, and Use Cases: A Review Paper

Yuru Lin[1], Krishnaveni Parvataneni[2]

[1]lyuru05@gmail.com

The Baldwin School
701 Montgomery Avenue, Bryn Mawr, Pennsylvania, USA 19010

[2]kveniparvataneni@gmail.com

BASIS Independent Silicon Valley
1290 Parkmoor Avenue, San Jose, California, USA 95126

Abstract— **As artificial intelligence (AI) becomes increasingly developed, technologies such as deepfake is becoming more advanced and widely used. Deepfake is a synthetic media creation process in which certain features of an existing image or video are replaced by the features of someone or something else. Generative Adversarial Networks (GAN) is generally considered to be the most effective method in creating deepfake media. Deepfake is useful in various fields, such as politics, translation, education, and entertainment. This paper reviews the generation, detection, and use cases of deepfake.**

Keywords— **deepfake, deepfake detection, GAN, deepfake uses, social psychology**

I. Introduction

Deepfake is a relatively new technology that was popularized on the internet in 2017. It uses artificial intelligence to automatically generate or alter images and videos. Its name originated from a Reddit user from the community r/Deepfakes, where people shared deepfake media they created. Deepfake is a form of media that involves a video made with an actor, and a person whose features are imposed over the actor's. It is made using autoencoders and Generative Adversarial Networks (GANs). An autoencoder is a type of neural network consisting of an encoder and a decoder. The encoder reduces the image to a lower-dimensional latent form, while the decoder reconstructs an image from the latent form. This representation is upgraded using a GAN which trains a discriminator with the decoder in an adversarial relationship, creating new images from the latent representation.

The most common detection method for images is looking for inconsistencies in their RGB values. This method is also useful for deepfake video detection. However, inspecting the inconsistency between frames produces better accuracy.

Deepfake could be used for various purposes in many different fields, including politics, translation, education, and entertainment. However, it could also be used for harm, with major examples being scam, generating false information, and creating pornography videos.

This review paper will explore the background of deepfake, technologies related to deepfake generation and detection, uses and use cases of deepfake, and regulations regarding deepfake.

II. Method

This paper reviews the subject of deepfake from the perspective of artificial intelligence (AI), along with psychology. The paper analyzes deepfake architecture, various uses of deepfake, and deepfake detection models.

Databases used when selecting reference papers are Science Direct, IEEE, and arXiv. News articles from publications including The Wall Street Journal, The Guardian, and The Washington Post were also referenced. Webpages, such as contents posted on the official website of organizations mentioned in this paper, and documentations from Congress, were used as well. The distribution of reference papers from different sources in shown in Fig. 1. Backward snowballing, the practice of finding articles from the citation of a study, is used. A total of 51 reference sources were selected.

There were two phases of research: a general landscape survey on Google and databases and a specific search in databases. The keywords used were fake news, deepfake, deepfake detection, deepfake image detection, deepfake video detection, deepfake uses, and deepfake history. Boolean operators were not heavily relied on.

As the this paper focuses on deepfake technology, all of the reference sources selected contain information related to the artificial intelligence discipline. Many of the sources also has a main focus on laws and regulations, politics, and social aspects. (Fig. 2). All of them were published post-2017 (Fig. 3. They were written in English, with most of them from the U.S. and the rest from English-speaking European countries such as the UK.

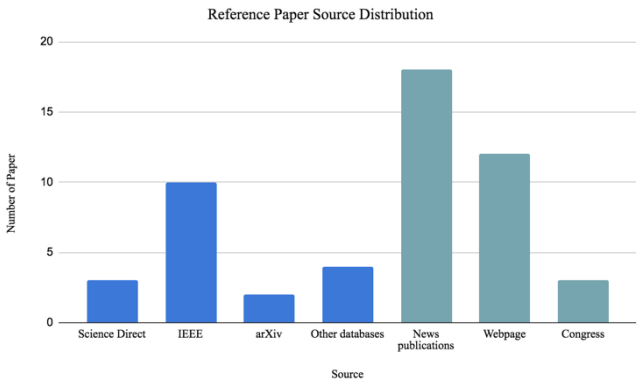

Fig. 1 Distribution of reference papers across databases

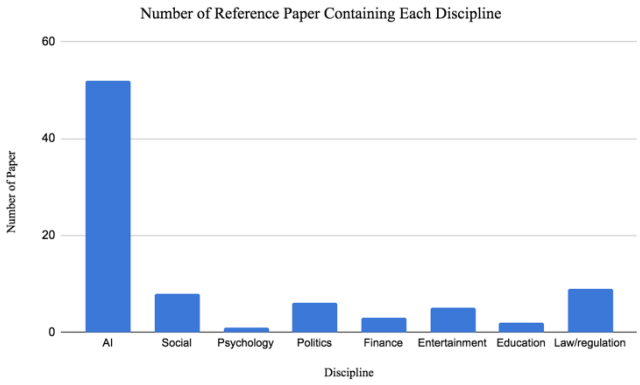

Fig. 2 Distribution of reference papers across disciplines

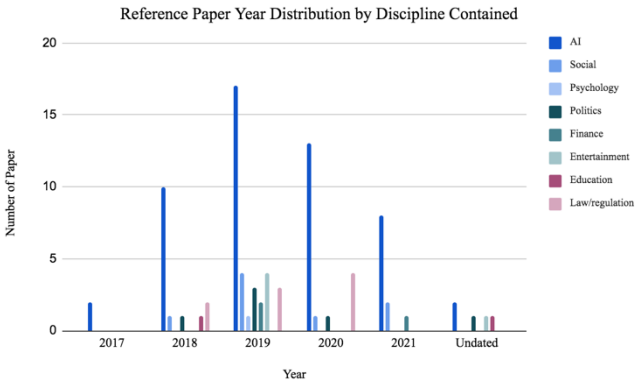

Fig. 3 Publication years of reference papers by discipline contained

III. Deepfake Background and Technology

A. Development of Deepfake

The first fully automated facial reanimation system was the Video Rewrite program published in 1997 by Cristoph Bregler, Michele Covell, and Malcolm Slaney [1]. However, deepfake was not popularized until 2017 when a Reddit user "u/deepfakes" posted pornography videos made by swapping female celebrities' faces onto existing videos [2], [3]. Even in 2019, pornography videos consist of 96% of deepfake videos online and have over 134 million views [3], [4]. This use of deepfake is a factor that contributes to its negative connotation.

Commercial development of deepfake, such as the publication of deepfake apps, rose since January 2018. Most of these apps were made for the purpose of entertainment. For instance, FakeApp was an app that allowed users to create and share videos with their faces swapped with each other [5]. By 2019, it was replaced by a relatively more open-source app called Faceswap. However, there were also some malicious apps. One was Deepnude, which could "take off" the clothes on a woman's picture. Fortunately, it was only up on windows and linux for about a month before it was taken down [6]–[8]. By 2020, apps such as Reface, Wombo, Avatarify, and MyHeritage allowed for more uses of deepfake. Reface allows the user to replace the face of a celebrity with their own. Avatarify allows the user to switch their own expressions onto someone else's face. While its videos are not as sophisticated as some others found on the internet, the app has been very popular. Similarly, Wombo

can turn a photo into a funny lip sync video. MyHeritage's "Deep Nostalgia" allows people to animate old photos of deceased loved ones and bring them to life [9].

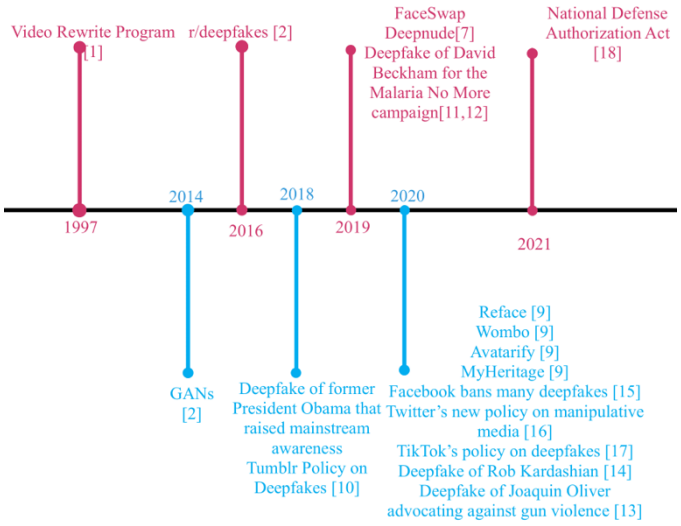

Fig. 4 Timeline of deepfake technology development [1], [2], [7], [9]–[18]

B. Deepfake Generation Methods

Generative Adversarial Networks (GAN) is generally considered to be the most effective method in creating deepfake media [2]. It consists of two parts: the generator and the discriminator. First, a deepfake image or video is generated in the generator, then the traces of deepfake are detected in the discriminator. The deepfake traces found in part two are then used as feedback to improve part one. When the discriminator is unable to detect the difference between real images and the synthetic ones from the generator, the deepfake generation is then successful [19]–[23]. A diagram of this model is shown in Fig. 5.

One instance of the application of GAN is creating extremely realistic-looking faces, such as the ones on the "This Person Does Not Exist" website [24]. There are many variations of GAN designed for different purposes; these variations include CYCLEGAN, STARGAN, ATTGAN, GDWCT, STYLEGAN, STYLEGAN2, PROGAN, FACEFORENSICS++, IMLE, and SPADE [25]–[27].

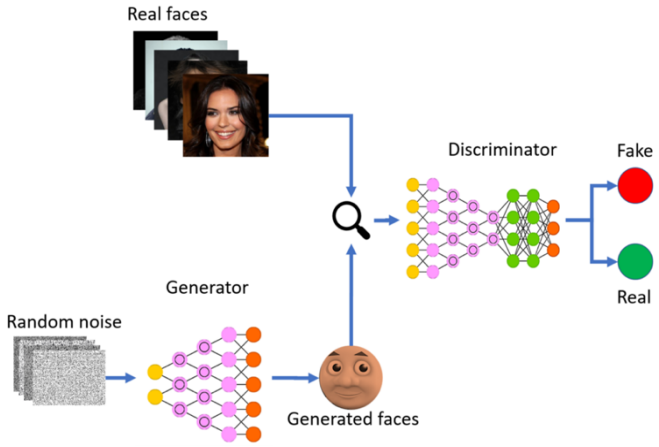

Fig. 5 GAN model diagram [28]

C. Deepfake Detection Methods

There are many detection methods for deepfake images and videos, some of which are shown in Table 1. The most common detection method for images is looking for inconsistencies in their RGB values. This method is also useful for deepfake video detection. However, inspecting the inconsistency between frames produces better accuracy; this inter-frame analysis technology can be seen in recent detection methods. A variety of neural networks can be used for deepfake detection. Convolutional neural networks (CNN) are proficient at comparing different frames. However, it is unable to compare the relationship between different pixels and thus cannot be used to detect deepfake images through the RGB method.

TABLE I
DEEPFAKE DETECTION METHODS COMPARISON

Year	Study Name	Detection Method	Result	Advantage	Limitation
2021	Learning Spatio-temporal features to detect manipulated	3-D CNN model that learns Spatio-temporal features from an adjacent frame sequence in a	> 99% accuracy on FaceForensics	- Analyzes inter-frame correlations	- Only specialized in detecting videos generated by FaceForensics++ and

	facial videos created by the Deepfake techniques [29]	video	++ and VidTIMIT datasets		VidTIMIT
2021	Deepfake Video Detection through Optical Flow based CNN [30]	Recognize motion dissimilarities in the temporal structure of a video by analyzing optical flow fields	81.61% accuracy on Face2Face	- Analyzes inter-frame correlations	- Cannot detect on images
2020	DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network [31]	VGG network analyzing RGB and noise characteristics of face images.	85.7 % accuracy	- Effectively detects deepfake images	- Cannot detect on videos
2020	DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern [32]	DeepVision algorithm observes blinking patterns of different demographics and compares them to the subject video.	87.5% accuracy	- Able to detect deepfake videos when its RGB and frames are consistent	- Subject video must involve humans - May be inapplicable to those with mental illnesses and nerve conduction pathway problems
2019	Image Feature Detectors for Deepfake Video Detection [33]	SVM classifier trained by feature points extracted from faces in deepfake videos.	94.5% accuracy with HOG feature extraction algorithm	- Effectively identify and extract feature points from images	- Used small dataset
2018	Deepfake Video Detection Using Recurrent Neural Networks [34]	Trains LSTM to classify deepfake videos with frame-level features extracted with CNN	97.1% accuracy with 80 frames	- Only needs a few seconds of a video	- Cannot detect on images

IV. Uses of Deepfake

A. Deepfake in Politics

Although deepfake is not currently widespread in generating political fake news, it is beginning to enter the political field [3]. For instance, in 2019, a deepfake video of former Italian prime minister Matteo Renzi portrays him making rude gestures toward other politicians by transferring his features onto the face of a comedian. Though many people were able to recognize that the video was altered from the absurdity of its content, those that did not were outraged [35], [36]. Considering that deepfake media is becoming increasingly difficult to discern as it develops, critics were alarmed at the potential damage that the technology can cause in politics. Some even believes it is a threat to democracy [2]. Additionally, they are also concerned that deepfake could be used as an excuse for politicians to dismiss politically damaging things that they actually did, a phenome known as the liar's dividend [37].

B. Deepfake in Finance

Deepfake could also be a threat to businesses. Its ability to alter the face or voice of an individual could be used to swindle funds. In 2019, a scammer altered his voice using deepfake and instructed the CEO of a U.K.-based energy firm to transfer €220,000 to a supposed Hungarian supplier. The CEO, believing the call was from his superior, complied and transferred the fund to the scammer. This instance was one of the first economic crimes that relied on deepfake [38], [39]. Deepfake could also be used for identity theft, cyber extortion, and stock manipulation through fabricated events [40].

C. Deepfake in Languages

Deepfake can be used to remove language barriers. Since deepfake can replicate voices and alter videos, it allows the actor of a video to speak various languages. Furthermore, it can make the voice to sound like the actor's and sync lip movements to words. For example, David Beckham, an English football player who is only fluent in one language, was shown to speak 9 different languages to share a message for a video in the Malaria Must Die Campaign [11], [12].

D. Deepfake in Animating the Dead

In education, deepfake could be used to bring historical figures to life and create more interactive classes. For instance, a team from the company CereProc used deepfake to analyze President John F. Kennedy's speeches to make a deepfake video of him reciting the speech that he was on his way to deliver when was assassinated [41]. The Illinois Holocaust and Museum Education Center used deepfake to create a showcase with the holograms of 15 Holocaust Survivors, and allowed visitors to interact with them and ask questions [42].

Deepfake could also resurrect the dead for other purposes than history education. One major instance was with Joaquin Oliver, a victim of the Parkland Shooting. His parents worked with McCann Health to produce a deepfake video that advocated for gun safety [13]. Another, albeit less important, example of such use of deepfake was when Kim Kardashian, an American media personality, worked with the company Kaleida to make and post a video of her late father. Additionally, companies such as MyHeritage can create realistic short videos of late family members by animating an old photo [14].

E. Deepfake in Entertainment

In entertainment, deepfake could help create virtual museums. For instance, Samsung AI Center and Skolkovo Institute of Science and Technology manipulated paintings such as Mona Lisa to make its eyes, head, and mouth move as though she was watching the audience. In another instance, the Dali Museum in Florida created moving portrait of Salvador Dali so that visitors could listen to him talk about himself [43]. Deepfake could also be used to generate video game characters. In the game *Un'emozione per sempre 2.0* created by Joseph Ayerle, deepfake was used to generate a character with the face of Ornella Muti, an Italian actress popular during the 1980s, and the body of Kendall Jenner, an American model [44].

Deepfake is also highly useful in the movie industry. In movies with flashbacks, the young version of a character is usually played by an actor different from the one that played the adult character. This could result in discrepancies in the appearance of the same character. This problem can be solved using deepfake. By using clips of actors in earlier stages of their life, deepfake, combined with computer graphics, can make actors look like younger versions of themselves [45]. This method was used for the resurrection of movie actor James Dean. He only had 3 major movie performances before his death in 1955 at the age of 24, but more than 60 years later, he can now star in a new film, named "Finding Jack", about the Vietnam War [46]. Despite the seemingly good outcome, this use of deepfake received significant criticism. It was deemed as "shameful" and disrespectful to Dean by the actor Chris Evans [47].

F. Deepfake in Pornography

As of 2019, pornography videos consist of 96% of deepfake videos online and have over 134 million views [4]. These videos are usually published without consent. The subjects of the videos are mostly female celebrities, with a quarter of them being South Korean stars [8]. In June of 2019, an app named Deepnude was released. It used GANs to remove clothing from women's pictures. Fortunately, it was taken down by June 27 of the same year. Additionally, another concern of deepfake in pornography is the use of deepfake for the production of child pornography [7].

V. Deepfake Policies and Regulations

A. Social Media Policies

In response to the spread of deepfake online, many social media companies have posed regulations restricting or banning deepfake media on their platforms.

An earlier policy regarding deepfake came from Tumblr's Community Guidelines in September 2018. While Tumblr's policies are not as strict on deepfake as other social media platforms, it banned certain types of deepfake pornography and non-consensual "creepshots". The policies were mostly related to sexualization and sexual harassment [48].

Facebook's deepfake regulations, Enforcing Against Manipulated Media, were announced in January of 2020 [15]. Since Instagram is owned by Facebook, it has the same policy regarding deepfake. Facebook and Instagram would remove any media produced by AI or machine learning that

"merges, replaces, or superimposes content onto a video, making it appear authentic", and could mislead people into believing something that did not actually happen. This policy excludes media edited for clarity or quality. It also doesn't extend to parodical or satirical content and videos that change the order of words or omit words.

Twitter's new policy on manipulative media, Synthetic and Manipulated Media Policy, was announced in February of 2020 [16]. It consists of 3 questions: "Is the content significantly and deceptively altered or fabricated?", "Is the content shared in a deceptive manner?", and "Is the content likely to impact public safety or cause serious harm?". Fig. 6 displays Twitter's responses to tweets containing deepfake contents based on the answers to the aforementioned questions depend. Labeling provides additional context on tweets, including a label to the content, a warning before content is shared or liked, reduced visibility of the content, or turning off likes, replies, and retweets. Tweets that meet all 3 criteria are subject to removal from the platform.

Is the content significantly and deceptively altered or fabricated?	Is the content shared in a deceptive manner?	Is the content likely to impact public safety or cause serious harm?	
✓	✗	✗	Content **may** be labeled.
✗	✓	✗	Content **may** be labeled.
✓	✗	✓	Content is **likely** to be labeled, or **may** be removed.*
✓	✓	✗	Content is **likely** to be labeled.
✓	✓	✓	Content is **likely** to be removed.

Fig. 6 Twitter's deepfake policy [16]

TikTok's policy on deepfake, Combating Misinformation and Election Interference on TikTok, came later than the other major social media platforms' [17]. In August 2020, while facing threats of bans in many countries including the US, TikTok explicitly banned deepfake contents, especially ones related to elections. While major political campaigns have not used deepfake, the Trump campaign used shallow fakes, which were less sophisticated edits that were similarly misleading. Though TikTok had already banned political ads, the deepfake ban, according to the company, makes spreading deceptive media for political gain more difficult.

B. Legislative Acts

Due to the potential issues posed by deepfake, several deepfake-related acts were posed. However, as a result of various problems, including difficulty in execution, a portion of these acts were unable to pass.

The US National Defense Authorization Act (NDAA) is a series of federal laws allocating the budget/expenditures of the Department of Defense. The first NDAA was passed in 1961. Since 2020, Congress has included provisions addressing the problem of deepfake as part of the NDAA. The 2021 NDAA required the Department of Homeland Security, for the next 5 years, to issue an annual report on deepfake, expanding the scope of the deepfake report called for in 2020's NDAA [18].

The DEEPFAKES Accountability Act, also known as the "Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019", was introduced by Congresswoman Yvette Clarke in 2019. The act was intended to restrict the damage done by synthetic media, partly by requiring people creating synthetic media to disclose that the media was altered. However, due to the lack of advanced deepfake detection models, it is difficult to determine whether an image or video is synthetic or not. This implies that it is almost impossible to enforce the DEEPFAKES Accountability Act. While this act never passed and was known as a stretch, it was a start to having legislation that would regulate deepfake media [49], [50].

The Identifying Outputs of Generative Adversarial Networks Act was proposed in December 2019, though it was never passed into law. This bill was intended to support research on deepfake generation models, such as GANs, and information authenticity [51], [52].

VI. Conclusions

Deepfake is a relatively new technology that was popularized on the internet in 2017. It uses artificial intelligence to automatically generate or alter images and videos. The term deepfake originated from a Reddit user from the community r/deepfakes, where people shared deepfake media they created. Deepfake is a form of media that involves a video

made with an actor, and a person whose features are imposed over the actor's. It is made using autoencoders and Generative Adversarial Networks (GANs). An autoencoder is a type of neural network consisting of an encoder and a decoder. The encoder reduces the image to a lower-dimensional latent form, while the decoder reconstructs an image from the latent form. This representation is upgraded using a GAN which trains a discriminator with the decoder in an adversarial relationship, creating new images from the latent representation.

The most common detection method for images is looking for inconsistencies in their RGB values. This method is also useful for deepfake video detection. However, inspecting the inconsistency between frames by examining spatio-temporal features produces better accuracy. Various neural network models, with a popular one being CNN, is fit for this purpose.

Deepfake is used for various purposes; some are negative, and others are positive. It could be used to generate fake or misleading news against political candidates, impersonate someone else in financial scams, or create nonconsensual pornography videos. On the other hand, it could also be used to animate historical figures in education, voiceover and lip-sync actors to various languages, and creating animations of a late loved one.

Due to the cases of deepfake being used to generate false information that was manipulatively spread, many social media companies have posed regulations restricting the use of deepfake of their platform. Tumblr, Facebook, Instagram, Twitter, and TikTok are among the major platforms that announced they had banned certain malevolent deepfake media and would remove if they appear.

In addition to social media regulations, there were several legislative acts in the U.S. regarding deepfake that were established or attempted to be established. These included The US National Defense Authorization Act (NDAA), The DEEPFAKES Accountability Act, and The Identifying Outputs of Generative Adversarial Networks Act.

ACKNOWLEDGMENT

The authors would like to thank Dr. Catherine Fang and the Apple-Pi institute for the facilities to do this research.

This research did not receive any specific grant from funding agencies in the commercial, public, or not-for-profit sectors.

REFERENCES

[1] "Video Rewrite, Origins of Deepfakes : History of Information," 2017. https://www.historyofinformation.com/detail.php?id=4792 (accessed Sep. 10, 2021).

[2] S. Adee, "What Are Deepfakes and How Are They Created?," *IEEE Spectrum*, Apr. 29, 2020. https://spectrum.ieee.org/what-is-deepfake (accessed Sep. 10, 2021).

[3] S. Adee, "World's First Deepfake Audit Counts Videos and Tools on the Open Web," *IEEE Spectrum*, Oct. 07, 2019. https://spectrum.ieee.org/the-worlds-first-audit-of-deepfake-videos-and-tools-on-the-open-web (accessed Sep. 10, 2021).

[4] "Overview and Key Findings of the 2020 Digital News Report," *Reuters Institute Digital News Report*, May 22, 2020. https://www.digitalnewsreport.org/survey/2020/overview-key-findings-2020/ (accessed Jul. 11, 2021).

[5] K. Reilly and S. Kovach, "Face-swapping videos could lead to more 'fake news,'" Apr. 13, 2018. https://www.businessinsider.com/fakeapp-lets-people-make-fake-videos-deepfakes-2018-4 (accessed Sep. 26, 2021).

[6] M. Masood, M. Nawaz, K. M. Malik, A. Javed, and A. Irtaza, "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward," 2021.

[7] S. Samuel, "A guy made a deepfake app to turn photos of women into nudes. It didn't go well.," *Vox*, Jun. 27, 2019. https://www.vox.com/2019/6/27/18761639/ai-deepfake-deepnude-app-nude-women-porn (accessed Sep. 26, 2021).

[8] H. Ajder, G. Patrini, F. Cavalli, and L. Cullen, "The State of Deepfakes: Landscape, Threats, and Impact," Sep. 2019, [Online]. Available: https://enough.org/objects/Deeptrace-the-State-of-Deepfakes-2019.pdf

[9] G. Fowler, "Anyone with an iPhone can now make deepfakes. We aren't ready for what happens next.," *Washington Post*, Mar. 25, 2021. Accessed: Sep. 26, 2021. [Online]. Available: https://www.washingtonpost.com/technology/2021/03/25/deepfake-video-apps/

[10] BuzzFeedVideo, *You Won't Believe What Obama Says In This Video!* 🤔, (Apr. 17, 2018). Accessed: Nov. 15, 2021. [Online Video]. Available: https://www.youtube.com/watch?v=cQ54GDm1eL0

[11] Malaria Must Die, *David Beckham speaks nine languages to launch Malaria Must Die Voice Petition*, (Apr. 09, 2019). Accessed: Sep. 10, 2021. [Online Video]. Available: https://www.youtube.com/watch?v=QiiSAvKJIHo

[12] O. Oakes, "'Deepfake' voice tech used for good in David Beckham malaria campaign," 2019. https://www.prweek.com/article/1581457?utm_source=website&utm_medium=social (accessed Sep. 10, 2021).

[13] "McCann Health New York - unfinished votes," 2020. https://www.mccannhealthnewyork.com/work/unfinished-votes/ (accessed Sep. 26, 2021).

[14] M. Dunne-Miles, "Deepfakes, dead relatives and digital resurrection," *The Face*, Apr. 06, 2021. https://theface.com/society/deepfakes-dead-relatives-deep-nostalgia-ai-digital-resurrection-kim-kardashian-rob-kardashian-grief-privacy (accessed Sep. 26, 2021).

[15] M. Bickert, "Enforcing Against Manipulated Media - About Facebook," Jan. 06, 2020. https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/ (accessed Sep. 10, 2021).

[16] "Our synthetic and manipulated media policy | Twitter Help," Feb. 04, 2020. https://help.twitter.com/en/rules-and-policies/manipulated-media (accessed Sep. 10, 2021).

[17] V. Pappas, "Combating misinformation and election interference on TikTok," *Newsroom | TikTok*, Aug. 05, 2020. https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok (accessed Sep. 10, 2021).

[18] N. F. DeBruyne and B. S. Torreon, "Defense Authorization and Appropriations Bills: FY1961-FY2018," Apr. 2018.

[19] T. T. Nguyen, Q. V. H. Nguyen, C. M. Nguyen, D. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep Learning for Deepfakes Creation and Detection: A Survey," *ArXiv190911573 Cs Eess*, Apr. 2021, Accessed: Sep. 16, 2021. [Online]. Available: http://arxiv.org/abs/1909.11573

[20] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, Jan. 2018, doi: 10.1109/MSP.2017.2765202.

[21] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, Oct. 2020, doi: 10.1145/3422622.

[22] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative adversarial networks: introduction and outlook," *IEEECAA J. Autom. Sin.*, vol. 4, no. 4, 2017, doi: 10.1109/JAS.2017.7510583.

[23] G. Developers, "Overview of GAN Structure | Generative Adversarial Networks," *Google Developers*, May 24, 2019. https://developers.google.com/machine-learning/gan/gan_structure (accessed Sep. 27, 2021).

[24] P. Wang, "This Person Does Not Exist," Feb. 2019. https://www.thispersondoesnotexist.com/ (accessed Sep. 10, 2021).

[25] L. Guarnera, O. Giudice, and S. Battiato, "Fighting Deepfake by Exposing the Convolutional Traces on Images," *IEEE Access*, vol. 8, pp. 165085–165098, 2020, doi: 10.1109/ACCESS.2020.3023037.

[26] T. Shen, R. Liu, J. Bai, and Z. Li, "'Deep Fakes' using Generative Adversarial Networks (GAN)," 2018.

[27] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 8789–8797. doi: 10.1109/CVPR.2018.00916.

[28] A. Deshpande, "What is GAN? The First Blog Post You Should Read — Simple Explanation for a Layman," *The Startup*, May 11, 2020. https://medium.com/swlh/a-simplistic-description-of-what-generative-adversarial-networks-actually-do-105a5ca2a729 (accessed Sep. 28, 2021).

[29] X. H. Nguyen, T. S. Tran, V. T. Le, K. D. Nguyen, and D.-T. Truong, "Learning Spatio-temporal features to detect manipulated facial videos created by the Deepfake techniques," *Forensic Sci. Int. Digit. Investig.*, vol. 36, Mar. 2021, doi: 10.1016/j.fsidi.2021.301108.

[30] R. Caldelli, L. Galteri, I. Amerini, and A. Del Bimbo, "Optical Flow based CNN for detection of unlearnt deepfake manipulations," *Pattern Recognit. Lett.*, vol. 146, Jun. 2021, doi: 10.1016/j.patrec.2021.03.005.

[31] X. Chang, J. Wu, T. Yang, and G. Feng, "DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network," in *2020 39th Chinese Control Conference (CCC)*, Jul. 2020, pp. 7252–7256. doi: 10.23919/CCC50068.2020.9189596.

[32] T. Jung, S. Kim, and K. Kim, "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," *IEEE Access*, vol. 8, pp. 83144–83154, 2020, doi: 10.1109/ACCESS.2020.2988660.

[33] F. F. Kharbat, T. Elamsy, A. Mahmoud, and R. Abdullah, "Image Feature Detectors for Deepfake Video Detection," in *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, Nov. 2019, pp. 1–4. doi: 10.1109/AICCSA47632.2019.9035360.

[34] D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," presented at the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Nov. 2018. doi: 10.1109/AVSS.2018.8639163.

[35] S. Venkataramakrishnan, "Can you believe your eyes? How deepfakes are coming for politics," *Financial Times*, Oct. 24, 2019. Accessed: Sep. 10, 2021. [Online]. Available: https://www.ft.com/content/4bf4277c-f527-11e9-a79c-bc9acae3b654

[36] "Deepfake video of former Italian PM Matteo Renzi sparks debate in Italy," *The Observers - France 24*, Oct. 08, 2019. https://observers.france24.com/en/20191008-deepfake-video-former-italian-pm-matteo-renzi-sparks-debate-italy (accessed Sep. 10, 2021).

[37] P. Chadwick, "The liar's dividend, and other challenges of deep-fake news," *the Guardian*, Jul. 22, 2018. http://www.theguardian.com/commentisfree/2018/jul/22/deep-fake-news-donald-trump-vladimir-putin (accessed Sep. 10, 2021).

[38] J. Damiani, "A Voice Deepfake Was Used To Scam A CEO Out Of $243,000," Sep. 03, 2019. https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=4902cecd2241 (accessed Sep. 10, 2021).

[39] C. Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case - WSJ," Aug. 30, 2019. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 (accessed Sep. 10, 2021).

[40] J. Bateman, "Deepfakes and Synthetic Media in the Financial System: Assessing Threat Scenarios," Carnegie Endowment for International Peace, 2020. Accessed: Sep. 10, 2021. [Online]. Available: https://www.jstor.org/stable/resrep25783.2

[41] "JFK Unsilenced | CereProc Text-to-Speech." https://www.cereproc.com/en/jfkunsilenced (accessed Sep. 26, 2021).

[42] E. Braunstein, "At this Holocaust museum, you can speak with holograms of survivors," Jan. 27, 2018. https://www.timesofisrael.com/at-this-holocaust-museum-you-can-speak-with-holograms-of-survivors/ (accessed Sep. 26, 2021).

[43] Y. Lu, "Deepfakes in Art and Museums CAFA Art Museum." https://www.cafamuseum.org/en/exhibit/newsdetail/2527 (accessed Sep. 26, 2021).

[44] J. Ayerle, "02-04-2018-the-italian-game-un-emozione-per-sempre-2-0 | ArtConnect," Mar. 2019. https://www.artconnect.com/projects/02-04-2018-the-italian-game-un-emozione-per-sempre-2-0 (accessed Sep. 26, 2021).

[45] T. Bradshaw, "Deepfakes: Hollywood's quest to create the perfect digital human," *Financial Times*, Oct. 10, 2019. Accessed: Sep. 27, 2021. [Online]. Available: https://www.ft.com/content/9df280dc-e9dd-11e9-a240-3b065ef5fc55

[46] A. Ritman and A. Ritman, "James Dean Reborn in CGI for Vietnam War Action-Drama (Exclusive)," *The Hollywood Reporter*, Nov. 06, 2019. https://www.hollywoodreporter.com/movies/movie-news/afm-james-dean-reborn-cgi-vietnam-war-action-drama-1252703/ (accessed Sep. 26, 2021).

[47] R. Pomeroy, "This iconic filmstar will star in a new movie - from beyond the grave," *World Economic Forum*, Nov. 08, 2019. https://www.weforum.org/agenda/2019/11/james-dean-cgi-deepfakes/ (accessed Sep. 26, 2021).

[48] Tumblr Staff, "Our Community Guidelines are changing," *Tumblr Staff*, Aug. 27, 2018. https://staff.tumblr.com/post/177449083750/new-community-guidelines (accessed Sep. 10, 2021).

[49] Y. D. Clarke, "Text - H.R.3230 - 116th Congress (2019-2020): Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019," Jun. 28, 2019. https://www.congress.gov/bill/116th-congress/house-bill/3230/text (accessed Sep. 10, 2021).

[50] D. Coldewey, "DEEPFAKES Accountability Act would impose unenforceable rules — but it's a start," *TechCrunch*, Jun. 13, 2019. https://social.techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/ (accessed Sep. 10, 2021).

[51] Congress, "S. Rept. 116-289 - IDENTIFYING OUTPUTS OF GENERATIVE ADVERSARIAL NETWORKS ACT," Nov. 09, 2020. https://www.congress.gov/congressional-report/116th-congress/senate-report/289 (accessed Sep. 10, 2021).

[52] M. F. Ferraro, "Deepfake Legislation: A Nationwide Survey," Sep. 2019.