The World of Deepfakes

Olivia Cabrera

Department of Journalism and Media Management, University of Miami

Abstract

This paper will review the major aspects of deepfake audio and video technology. First, it will provide an in-depth explanation of how deepfakes are created using artificial intelligence. Second, the paper will focus on key applications of deepfakes: on social media, in politics and Hollywood, in art, and in the healthcare industry. As the number of deepfakes online continues to grow, the number of advantages and disadvantages of the technology also increases. The last sections of the paper will explain the laws surrounding deepfakes in the United States and how a media consumer can detect deepfake content. This paper concludes that the current utilizations of deepfakes are more harmful to society than beneficial and the technology of deepfake audio and video will continue to improve.

Keywords: deepfake, artificial intelligence, media, video, audio, technology

The World of Deepfakes

The purpose of this research is to examine deepfakes, how they work, and their effects on society. Deepfakes refer to artificial intelligence (AI) generated characters that feature a human appearance and seem authentic to the naked eye. Most commonly, deepfakes utilize face-swapping technology in which a video of a person's face is replaced with the likeness of someone else. Deepfake technology is often associated with media misinformation and manipulation of the public; however, the technology is also being used in beneficial ways in the healthcare and education industries.

Deepfakes are essential to understand because as technology advances, deepfakes will become even more realistic. As noted by *Forbes*, "deepfake technologies will play an increasingly constructive role in recreating the past and in envisioning future possibilities" (Chandler, 2020, para. 4). In December of 2018, a study by Deeptrace Labs revealed there were only 7,964 deepfake videos online (Ajder et al., 2019). Since the beginning of 2019, the number of deepfake videos has doubled every six months, leading to 85,000 deepfake videos online in December of 2020 (Petkauskas, 2021). Thus, this research will also raise awareness of deepfakes and their potential applications. It is crucial for society, more specifically social media users, to understand the dangers of misinformation that come with deepfakes.

This research sets out to address the topics of the technology, applications, concerns, and legalities of deepfakes. To begin, this paper will first examine how deepfakes are created with AI technology. The research will then explore the many applications of deepfakes, ways to recognize them, and advantages versus disadvantages. A conclusion that includes analysis and discussion will be presented after the description of deepfakes.

Technology Overview

Deepfake is a word that was created by combining the terms “deep learning” and “fake.” It is the name used to describe video and audio that utilize “artificial intelligence and algorithms to superimpose the actions and speech of one person onto another” (James, 2019, para. 2). The technology used to create these video and audio clips is constantly evolving, and there are currently several different ways to generate deepfakes. Every method of creating deepfakes requires a large amount of video, image, and audio data to produce a realistic outcome.

Technology of Video Deepfakes

The most common way to create deepfakes is through deep learning, a subset of machine learning, and involves the use of deep neural networks and autoencoders. *Business Insider* explains: “the autoencoder is a deep learning AI program tasked with studying video clips to understand what the person looks like from a variety of angles and environmental conditions” (Johnson, 2022, para. 12). The autoencoder uses these video clips in order to map one person’s face onto the face of another person in a video. Generative Adversarial Networks (GANs) represent a machine learning model that is used in the process of creating deepfakes. GANs detect flaws in deepfake videos and improve them, making deepfakes almost impossible to detect. In order to create realistic deepfakes, GANs use two neural networks to compete against each other. One of them is a generative neural network, which creates a realistic image of a person through decoding. The other neural network used is a discriminative classifier, which aims to correctly classify whether the image from the neural network is real or fake (Nguyen et al., 2022, p. 3). These two neural networks train together to improve their capabilities and produce extremely realistic images, while the discriminative classifier can accurately distinguish between real and fake.

There is a difference in how deepfakes are made when it comes to deepfake generators, such as apps and software programs. Deepfake generators are created using an encoder-decoder pair. In this method, an autoencoder extracts features of face images, and the decoder is then used to reconstruct the images. Two encoder-decoder pairs are needed to swap faces between source images and target images, where each pair is used to train on an image set and the encoder's parameters are shared between two network pairs. Nguyen et al. (2022) explain: "this strategy enables the common encoder to find and learn the similarity between two sets of face images, which are relatively unchallenging because faces normally have similar features such as eyes, noses, mouth positions" (p. 3). A deepfake creation model can be seen in Figure 1, where two neural networks are using the same encoder but different decoders (A and B) for training. A picture of the original face A is encoded with decoder B in order to create the deepfake, which can be seen at the bottom. This approach to creating deepfakes is used on platforms such as DeepFaceLab on GitHub. GitHub, an Internet hosting company for software development, has a large amount of deepfake software, making it cheap and easy for users to generate their own deepfake videos (GitHub, Inc., 2022). DeepFaceLab is the most popular software for creating deepfakes on GitHub. There is also a variety of apps available for download, such as Reface and Wombo, that make creating deepfakes easy for beginners. Reface is an app powered by GANs and has the ability to map a person's face onto an image or video of someone else, creating very realistic results (Maring, 2022). Users can choose from a variety of GIFs, film clips, and music videos and swap their faces with the face of the person in the clip or GIF. Wombo is an app that launched in 2021 and is an AI-based lip-syncing app (AWS, 2022). The user has to upload a photo of their face, choose a song, and Wombo will instantly turn the picture into a lip-syncing

video. While the artificial intelligence technology that provides the foundation for deepfakes involves many aspects, these apps allow for simple deepfake generation at the touch of a button.

Technology of Audio Deepfakes

The creation of deepfake audio is based on similar technology used by deepfake videos. There are three categories of deepfake audio: imitation-based, synthetic-based or text-to-speech (TTS), and replay-based (Almutairi et al., 2022). As seen in Figure 2, imitation-based deepfakes require an original and target audio to be recorded. In this type of audio recreation, the signal of the original audio will say the speech in the target audio "...using an imitation generation method that will generate a new speech" (Almutairi et al., 2022, p. 3). With this method, it is difficult for people to differentiate between the real and fake audio created. In synthetic-based or TTS, text is transformed into speech and consists of three parts: a text analysis model, an acoustic model, and a vocoder. "First, clean and structured raw audio should be collected, with a transcript text of the audio speech. Second, the TTS model must be trained using the collected data to build a synthetic audio generation model" (Almutairi et al., 2022, p. 3). There are a few generation models for synthetic-based audio, some of which include Deep Voice 3, Tactoran 2, and FastSpeech2. "In the synthetic technique, the transcript text with the voice of the target speaker will be fed into the generation model. The text analysis model then processes the incoming text and converts it into linguistic characteristics. Then, the acoustic module extracts the parameters of the target speaker from the dataset depending on the linguistic features generated from the text analysis module" (Almutairi et al., 2022, p. 3). Following this process, the vocoder will create speech waveform models based on the acoustic parameters and will generate the final audio file. The process of TTS can be seen in Figure 3. The final type of deepfake audio is called replay-based, also known as replay attacks. Replay attacks aim to replay a recording of the target

speaker's work and are used maliciously. The two types of replay-based attacks are far-field detection and cut-and-paste detection. "In far-field detection, a microphone recording of the victim recording is played as a test segment on a telephone handset with a loudspeaker. Meanwhile, cutting and pasting involves faking the sentence required by a text-dependent system" (Almutairi et al., 2022, p. 4). While each type of audio generation serves a different purpose, they are all considered to be forms of deepfake audio.

Overview of Applications

In 1997, research scientists Christoph Bregler, Michele Covell, and Malcolm Slaney invented a program called Video Rewrite. Video Rewrite took existing video footage of a person speaking and modified it to depict the person mouthing the words contained in a different audio track (Norman, 2022). This program was the first of its kind and relied on machine learning to automate facial reanimation. Video Rewrite could be considered the very first deepfake; however, the term "deepfake" itself was not used until 20 years later. The term was first utilized by a Reddit user in 2017 to describe a series of pornographic videos that used face-swapping technology to feature celebrities (Somers, 2020). While the Reddit videos were the first time deepfake videos had been posted online, this instance was only the beginning of people discovering the many ways deepfakes could be used.

Social Media, Celebrities, and Politicians

One of the most popular uses of deepfakes is on social media. Social media apps, such as TikTok and Instagram, feature deepfake videos, and a search of the hashtag "deepfake" will reveal thousands of results on both platforms. While some of these videos are deepfakes that users have made with simple apps such as Reface, most deepfakes on social media feature celebrities and politicians. There are accounts devoted to posting deepfakes, and Hollywood is a

large target of these videos. A TikTok account called @deeptomcruise is dedicated to creating parody videos of actor Tom Cruise. The videos show an extremely realistic-looking Tom Cruise singing, playing guitar, and hanging out with other celebrities, including Paris Hilton and Keegan-Michael Key. The account is home to almost four million followers, and the videos have a collective 14 million likes. There are also deepfake TikTok accounts dedicated to actor Keanu Reeves and actress Margot Robbie, which go by the names @unreal_keanu and @unreal_margot. Artificial intelligence deepfake technology also helped actor Val Kilmer get his voice back after losing it from surgery for throat cancer in 2015. A U.K.-based software firm that clones voices for actors, Sonantic, was able to digitally restore Kilmer's voice. "The generated voices have gotten more realistic in the age of deepfakes, a technology that uses AI to manipulate content to look and sound deceptively real" (Brown, 2021, para. 4). Sonantic used old recordings of Kilmer to recreate his voice and a video was posted on YouTube to reveal the actor's new AI voice. The software company created 40 versions of Kilmer's voice using existing audio, and "they created a script based on the material, linked the audio and text together in "short chunks" and ran the data through their "voice engine" algorithm, which learn to speak by listening to the recordings…" (Brown, 2021, para. 13). In order to use this technology, Kilmer had a desktop application where he could type anything he wanted to say into the model and alter the pitch and delivery (Mack, 2021). "As in Kilmer's case, the technology has possibilities for people who have difficulty speaking or actors needing to rest their vocal cords after long screaming sessions in the studio" (Brown, 2021, para. 18). While this instance of deepfake audio was beneficial to Kilmer, the ethics of the technology were called into question when director Morgan Neville used deepfake audio in a documentary about the late chef Anthony Bourdain. The documentary, *Roadrunner: A Film About Anthony Bourdain,* chronicles the life of Bourdain

but uses AI software to generate his voice for three lines of the film without disclosing it to viewers. In an interview with *GQ*, Neville discussed the synthetic audio, and Bourdain's fans were angered. "Neville used the A.I.-generated audio only to narrate the text that Bourdain himself had written" (Rosner, 2021, para. 2). Viewers argued that Bourdain should have been able to control how his words were delivered, and Bourdain's ex-wife Ottavia Bourdain was also upset by the use of this technology, claiming she never gave Neville approval to deepfake Bourdain's voice (Yang, 2021).

Politicians are another target for deepfakes, and videos featuring former Presidents Barack Obama and Donald Trump have garnered millions of views on YouTube. In the fall of 2022, a video of Vice President Kamala Harris from 2021 resurfaced where she appeared to say that all people hospitalized with Covid-19 were vaccinated (Hsu, 2022). The audio and video were altered, and the vice president had previously said that all people hospitalized were unvaccinated. Another altered TikTok posted in October of 2022 depicted President Biden singing the children's song "Baby Shark" instead of the national anthem. During the 2020 election, TikTok vowed to remove manipulated content, but there are still hundreds of fake videos of politicians on the platform. This content is threatening to politics because deepfake audio and video clips are becoming harder to distinguish from reality, and deepfake content could allow political misrepresentations to govern social media.

Pornographic Content

While most of the deepfakes on social media and the Internet are lighthearted and entertaining, many of them promote fabricated adult films. These deepfakes portray nonconsensual revenge pornography and often feature celebrities and politicians. There are websites dedicated to deepfake pornography, with the most prominent website being

MrDeepFakes. The founder of the website remains anonymous, but he claims his sole motivations to be his commitment to free speech and a desire to advance machine learning (Serwotka, 2022). The website has over 20,000 deepfake videos of female celebrities, with 25 being added daily by deepfake pornography producers. Despite these women being unable to consent to their faces being in the videos, the website is extremely profitable and was founded after deepfake adult content was banned from Reddit in 2018. *MrDeepFakes* is only one website out of many for deepfake pornography, and this type of content is on the rise. Fraud detection company Sensity AI conducted research in 2018 that revealed over 90% of deepfakes are nonconsensual pornographic videos of women, and search interest in deepfake pornography increased by 31% from 2021 to 2022 (Savin, 2022).

Other Applications

At the Dalí Museum in St. Petersburg, Florida, Salvador Dalí has come back to life through a deepfake. When visitors press the doorbell on the screen at the museum, Dalí tells them stories about his life and offers to take selfies with them. “The exhibition, called Dalí Lives, was made in collaboration with the ad agency Goodby, Silverstein & Partners (GS&P), which made a life-size recreation of Dalí using the machine learning-powered video editing technique” (Lee, 2019, para 2). GS&P used footage from interviews and machine learning to train the artificial intelligence algorithm on Dalí’s face. Dalí’s facial expressions were then imposed on an actor with Dalí’s body proportions, making this experience life-size. In order to master his voice, quotes from interviews were synced with a voice actor who could mimic Dalí’s accent. While there is ethical debate surrounding the topic of bringing the deceased back to life through technology, Dalí has no living family and the sole heir in his will is the Spanish Kingdom. The exhibition was run with permission from the Dalí Foundation in Spain. The purpose of Dalí

Lives is purely educational, and the experience "aims to have visitors empathize with Dalí as a human being" (Lee, 2019, para. 4). This educational application of bringing a historical figure back to life with a deepfake is just one example of the ways deepfakes are used outside of social media. In 2019, deepfake audio technology was used to scam the CEO of a UK-based energy company for $243,000 (Damiani, 2019). The CEO thought he was on the phone with his boss, the chief executive, when he "followed orders to immediately transfer €220,000 (approx. $243,000) to the bank account of a Hungarian supplier" (Damiani, 2019, para. 2). The fraudulent phone call used AI technology to mimic the chief executive's voice. The software used is unknown due to the fact that no suspects have been identified and this was the first public instance of deepfake audio being used for fraudulent purposes.

Advantages and Disadvantages of Deepfakes

In order to fully understand the applications of deepfake audio and video, it is crucial to examine both the advantages and disadvantages of deepfakes. The benefits of deepfakes include being used for educational purposes, in the healthcare industry, and in entertainment. Although there is a positive side to deepfakes, media disinformation and manipulation are only two of the many disadvantages of this technology.

Advantages

As previously mentioned, Val Kilmer has been given a second chance to speak thanks to deepfake technology. Deepfake audio has the ability to give people who are unable to speak a second chance to express themselves. In the healthcare industry, deepfakes can be used to make electronic medical records (EMRs) and protected health information (PHI) anonymous. As telemedicine becomes more universally adopted, patients may need to be recorded during appointments by medical professionals in order to detect symptoms for early disease prognosis

(Chen et al., 2021). "A GAN architecture similar to that used by the software faceswap can be used to de-identify faces in live videos, and similar methods could be used for the de-identification of EMRs, medical images, and other PHI" (Chen et al., 2021, para. 13). This adoption of deepfakes in the healthcare industry could prove beneficial in keeping patient's identities and data private. Deepfakes could also be of service to the advertising industry and be adopted to create lower-cost video campaigns. "Authorized deepfakes could allow marketers to feature huge stars in ads without requiring them to actually appear on-set before cameras, bringing down costs and opening new creative possibilities" (Coffee, 2022, para. 7). Deepfake technology also has the power to recreate things that no longer exist, and the aforementioned Dalí Lives exhibit is just one instance of this opportunity. In a classroom setting, for example, students now have the ability to hear former President John F. Kennedy's speech on the resolution of the Cold War. Using deepfake audio, *The Times* and creative agency Rothco created the JFK Unsilenced campaign (Collins, 2018). This campaign created audio of JFK giving the speech he was set to deliver on the day he was assassinated. Deepfakes have the power to allow for more engaging lessons in the classroom while also serving as entertainment.

Disadvantages

Many applications of deepfakes do not only cause concern to celebrities and politicians but also to the general public. Deepfakes serve as a form of disinformation to the public, which can be seen in the creation of political deepfakes that have the ability to pose a threat to democracy. Fake videos of political candidates could potentially sway election results and ruin reputations. In a discussion about deepfakes, United States Senator Marco Rubio stated: "Today… all you need is the ability to produce a very realistic fake video that could undermine our elections, that could throw our country into tremendous crisis internally and weaken us

deeply" (Toews, 2020, para. 18). As the amount of deepfake audio and video online grows, it is more difficult to have them identified and removed quickly. "In AI circles, reports *The Washington Post*'s Drew Harwell, identifying fake media has long received less attention, funding, and institutional support than creating it" (Galston, 2020, para. 7). The major disadvantage to deepfakes is the potential for deception, and deepfakes on the Internet can cause people to create opinions based on false information. The Brookings Institution explains that deepfakes are capable of "distorting democratic discourse; manipulating elections; … undermining public safety; and inflicting hard-to-repair damage on the reputation of prominent individuals…" (Galston, 2020, para. 11). Deepfakes also pose a threat to the journalism industry because mistakenly reporting on manipulated content has the ability to hurt the reputation and credibility of the journalist, victim of the deepfake, and profession as a whole (Morris, 2019). As previously discussed, deepfake audio has the power to generate fraudulent phone calls and could result in more instances similar to the scamming of the UK energy company. In August of 2022, technology company VMware, Inc. released its eighth annual Global Incident Response Threat Report, in which it stated that 66% of cybersecurity professionals had seen deepfakes used in a cyberattack (Kepczyk, 2022). Deepfakes can be used to impersonate people, steal identities, and make them targets of cybercrime.

The ethical issue of using deepfakes for nonconsensual sexually explicit content is also a huge disadvantage of the technology. Deepfakes could destroy the reputation of a victim who was digitally altered to be in a pornographic video. Pornographic content made with deepfakes does not only target celebrities. For instance, Helen Mort, a 36-year-old poet from England, still has nightmares about the time she discovered non-sexual images of her had been uploaded to an adult film website (Royle, 2021). Photos of her were taken from social media, and "Users on the

site were invited to edit the photos, merging Helen's face with explicit and violent sexual images" (Royle, 2021, para. 3). Mort was alerted to the deepfake content by an acquaintance and took the content to the police, but was told that nothing could be done. Mort mentioned, "Your initial response is shame and fear. I didn't want to leave the house. I remember walking down the street, not able to meet anyone's eyes, convinced everyone had seen it…" (Savin, 2022, para. 20). This case is just one of many instances where normal people have been maliciously targeted by deepfake pornography. Deepfake adult content is easily accessible online and creators take requests from viewers to deepfake their favorite celebrities, politicians, co-workers, and friends. Miles Fisher, the owner of the TikTok account @deeptomcruise, creates deepfakes of Tom Cruise without the actor's consent (Stump, 2021). While Fisher states he does not have any intentions to generate inappropriate content of the actor, the ethics of the content are up for debate due to the realistic nature of the videos. The defamatory use of deepfakes is harmful to society and the main disadvantage of the AI technology.

Deepfake Laws

Currently, the only legislation in the United States concerning deepfakes exists in Virginia, Texas, and California. In Virginia, the selling and distribution of deepfake pornography are outlawed (Loomis, 2022). The law in California prevents both the creation and distribution of deepfake pornography and "prohibits a person or other entity from distributing with "actual malice" doctored and otherwise deceptive material depicting a political candidate within sixty days of an election…" unless it has been disclosed to the public that the content has been manipulated (Inglesh, 2020, para. 11). In Texas, it is illegal to create and distribute a deepfake with intent to influence the results of an election within 30 days of the election (Artz, 2019).

While there are laws against revenge pornography in 46 U.S. states, only those in Virginia and California include deepfake pornography.

In 2019, The Deep Fakes Accountability Act was introduced to the U.S. House of Representatives (H.R.3230, 2019). The bill has not yet been passed but would require deepfake creators to comply with watermark and disclosure requirements. New criminal offenses related to the production of deepfakes that do not comply with requirements would also be established with The Deep Fakes Accountability Act. The Deepfake Task Force Act was introduced to the U.S. Senate in May of 2022, which would require the establishment of the National Deepfake Provenance Task Force (S.2559, 2022). "Comprised of federal and nonfederal stakeholders, the task force must address threats posed by digital forgeries (i.e., digital audio, images, and text fabricated or manipulated using artificial intelligence, machine learning, and other emerging technologies with the intent to mislead" (S.2559, 2022, para. 2). The task force would then have to develop a plan to reduce the growth of deepfakes and similar content.

Within the United States Department of Defense, The Defense Advanced Research Projects Agency (DARPA) has two programs committed to the detection of deepfakes: Media Forensics (MediFor) and Semantic Forensics (SemaFor). MediFor "was to develop algorithms to automatically assess the integrity of photos and videos and to provide analysts with information about how counterfeit content was generated" (Sayler et al., 2022, para. 12). Semafor technologies will build upon Medifor technologies and will "develop algorithms that will automatically detect, attribute, and characterize (i.e., identify as either benign or malicious) various types of deep fakes" (Sayler et al., 2022, para. 13). SemaFor will also work to "...prioritize suspected deepfakes for human review" (Sayler et al., 2022, para. 13).

Deepfake Detection

Facebook is one company that has already revealed it has come up with artificial intelligence technology to detect and track deepfakes on the app using reverse engineering (Diaz, 2021). “Facebook’s new software runs deepfake images through its network. Their AI program looks for cracks left behind in the manufacturing process used to change an image’s digital “fingerprint” (Diaz, 2021, para. 7). DeepMedia, a company that creates and detects deepfakes, has announced their plans to release a deepfake detection product to the public by December 2022 or early 2023 (Field, 2022). Users will be able to pay to upload content, “and receive a report examining whether the content is falsified, the algorithm that was initially used to create the deepfake, and how the company came to that conclusion” (Field, 2022, para. 5). Co-founder of DeepMedia, Rijul Gupta, claims that the company’s deepfake detectors currently work at 95% accuracy, but will not be released to the public until the detectors are at 99% accuracy. To check how well ordinary people can detect a deepfake, the Massachusetts Institute of Technology (MIT) created a website called *Detect Fakes* (Groh, 2022). The website shows a series of 32 videos, audio, and text, and the viewer has to decide whether or not the content was fabricated (Groh, 2022). *Detect Fakes* was designed by MIT to answer the question of how realistic deepfakes are and “identify techniques to counteract AI-generated misinformation” (Groh, 2022, para. 1).

Until deepfake detectors are widely used by the public, there are some key signs a media consumer can look for when trying to figure out if a video is real or fake. MIT’s Media Lab offers a few ways to check for deepfakes, and they all have to do with the face of the person in the video. First, the viewer should look at the eyes and eyebrows of the person because deepfakes often fail to accurately represent face shadows (Groh, 2022). If the person in the target video is wearing glasses, the viewer should look at the glare and see if the angle of the glare

changes when the person moves because deepfakes often fail to properly portray lighting. Another one of the main ways to tell if a video is a deepfake is to watch how the person blinks. Blinking is hard to reproduce in deepfake videos because, "with an average rate of 4.5 blinks per second and each blink lasting 0.1-0.4 seconds, most training datasets of videos used for deepfake detection have a scarce amount of faces with their eyes closed" (Pishori et al., 2022, p. 2). A lack of eye-blinking is a simple indicator that the video could be a deepfake. Some deepfakes are not as advanced as others and therefore are easier to spot. Most deepfakes are facial transformations and the best way to look for one is to closely analyze the face of the person in the video.

Conclusion

As artificial intelligence technology improves, there is no doubt that deepfakes will become even more realistic. An online search for deepfakes generates millions of results, most of which include websites to create deepfakes and places to watch deepfake content. Although deepfake technology is relatively new, it has had a tremendous impact on society and the media thus far. Deepfake accounts have millions of views on social media, and new content is being uploaded to these accounts every day, with people in the comments often questioning if the video is real or fake.

While there are some benefits to deepfakes and how they are used, the technology will continue to be taken advantage of until there are more laws regulating deepfake content. Deepfakes have the ability to easily distort reality and disinform viewers. I am confident that the number of malicious applications of deepfakes outweighs the positive ones, but I also see how this technology is beneficial to the education and art industries.

Deepfakes have the opportunity to pave the way toward increased media literacy among consumers. The number of deepfake detection agencies is growing, and these agencies along

with raising awareness of the technology can encourage viewers not to believe everything they see online. I believe it is crucial to teach media consumers about the dangers of deepfakes and how they have the power to influence beliefs, opinions, and even elections.

Laws in the United States are currently failing to protect victims of nonconsensual deepfake pornography. These videos have the ability to cause irreversible damage to peoples' reputations and threaten media consumers' capacity to receive information online. It can be argued that the desire for laws against deepfakes interferes with the First Amendment right to freedom of expression because not all uses of the technology are dangerous. However, the malicious applications of deepfakes could jeopardize many industries, which is why I think the technology needs to be regulated. There should be laws in all 50 states similar to those in California, Texas, and Virginia, that work to protect people targeted by defamatory deepfakes. The videos are often indistinguishable from reality and are extremely easy to create. On the other hand, exhibitions such as Dalí Lives are able to facilitate new experiences and can bring historical figures back to life for the purpose of education. The threat of deepfakes does not come from the technology used to create them but from the creators who have the intent to distort reality.

The advanced artificial intelligence technology used to create deepfake audio and video provides opportunities for more research. Technology companies now have the ability to explore the potential of deepfakes and improve on ways to detect them. While there have already been many instances of deepfakes being used maliciously, I hope that the technology will become used in increasingly beneficial ways. Deepfakes may be around forever, which is why it is necessary to understand how they are created and the implications of the technology.

References

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019, September). The state of deepfakes: Landscape, threats, and impact. *Deeptrace Labs.* https://enough.org/objects/Deeptrace-the-State-of-Deepfakes-2019.pdf

Almutairi, Z., & Elgibreen, H. (2022, May 4). A review of modern audio deepfake detection methods: Challenges and future directions. *Algorithms* 2022, *15,* pp.155. https://doi.org/10.3390/a15050155

Artz, K. (2019, October 11). Texas outlaws 'Deepfakes' – but the legal system may not be able to stop them. *Law.com.* https://www.law.com/texaslawyer/2019/10/11/texas-outlaws-deepfakes-but-the-legal-system-may-not-be-able-to-stop-them

Brown, D. (2021, August 18). AI gave Val Kilmer his voice back. But critics worry the technology could be misused. *The Washington Post.* https://www.washingtonpost.com/technology/2021/08/18/val-kilmer-ai-voice-cloning/

Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K., & Mahmood, F. (2021, June 15). Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng* 5, pp.493-497. https://doi.org/10.1038/s41551-021-00751-8

Coffee, P. (2022, October 25). 'Deepfakes' of celebrities have begun appearing in ads, with or without their permission. *The Wall Street Journal.* https://www.wsj.com/articles/deepfakes-of-celebrities-have-begun-appearing-in-ads-with-or-without-their-permission-11666692003

Collins, P. (2018, March 16). Kennedy's powerful speech still resonates today. *The Times.* https://www.thetimes.co.uk/article/kennedys-powerful-speech-would-still-resonate-today-kbxpqrnrl

Diaz, J. (2021, June 17). Facebook researchers say they can detect deepfakes and where they came from. *NPR.* https://www.npr.org/2021/06/17/1007472092/facebook-researchers-say-they-can-detect-deepfakes-and-where-they-came-from

Editorial Team, AWS. (2022, January 28). How AWS supported WOMBO's wildly popular, AI-powered app. *AWS.* https://aws.amazon.com/blogs/startups/how-aws-supported-wombos-wildly-popular-ai-powered-app/

Field, H. (2022, August 22). Meet the company working with the Air Force to detect deepfakes. *Emerging Tech Brew.* https://www.emergingtechbrew.com/stories/2022/08/22/meet-the-company-working-with-the-air-force-to-detect-deepfakes

Galston, W. A. (2020, January 8). Is seeing still believing? The deepfake challenge to truth in politics. *The Brookings Institute.* https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/

Github, Inc. (2022). https://github.com/

Groh, M. (2022). Detect DeepFakes: How to counteract misinformation created by AI. *MIT Media Lab.* https://www.media.mit.edu/projects/detect-fakes/overview/

H.R.3230 - 116th Congress (2019-2020): DEEP FAKES Accountability Act. (2019, June 28). https://www.congress.gov/bill/116th-congress/house-bill/3230

Hsu, T. (2022, November 4). Worries grow that TikTok is new home for manipulated video and photos. *The New York Times.* https://www.nytimes.com/2022/11/04/technology/tiktok-deepfakes-disinformation.html

Inglesh, A. (2020). Walking a fine line: Finding harmony between California's deep fake laws and the First Amendment. *Communications Lawyer, 35*(3), pp.27-31.

James, S. B. (2019, July 31). The business case for protecting the technology behind 'deepfakes'. *S&P Global Market Intelligence.* https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/the-business-case-for-protecting-the-technology-behind-deepfakes-53069518

Johnson, D. (2022, August 10). What is a deepfake? everything you need to know about the AI-powered fake media. *Business Insider*. https://www.businessinsider.com/guides/tech/what-is-deepfake

Kepczyk, R. H. (2022, October). Deepfakes emerge as real cybersecurity threat. *American Institute of Certified Public Accountants.* https://www.aicpa.org/news/article/deepfakes-emerge-as-real-cybersecurity-threat

Lee, D. (2019, May 10). Deepfake Salvador Dalí takes selfies with museum visitors. *The Verge.* https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum

Loomis, A. (2022, April 20). Deepfakes and American Law. *Davis Political Review.* https://www.davispoliticalreview.com/article/deepfakes-and-american-law

Mack, E. (2021, August 19). Hear Val Kilmer's voice, re-created by AI after throat cancer took it away. *CNET.*

https://www.cnet.com/culture/entertainment/hear-val-kilmers-voice-re-created-by-ai-after-throat-cancer-took-it-away/

Maring, J. (2022, March 3). Is Reface app safe? Here's what you should know before downloading. *Screenrant.* https://screenrant.com/is-reface-app-safe-use-privacy-concerns

Morris, A. (2019). Going Deep: In a world of "fake news" accusations, deepfakes may soon be a very real problem for journalists. *Quill, 107*(2), pp.21-25.

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q. V., & Nguyen, C. M. (2022, August 11). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding, Volume 223*, 2022, 103525. https://arxiv.org/pdf/1909.11573.pdf

Norman, J. (2022, October 22). Video Rewrite, origins of deepfakes. *HistoryofInformation.com.* https://www.historyofinformation.com/detail.php?id=4792

Petkauskas, V. (2021, September 28). Report: number of expert-crafted video deepfakes double every six months. *Cybernews.* https://cybernews.com/privacy/report-number-of-expert-crafted-video-deepfakes-double-every-six-months/

Pishori, A., Rollins, B., van Houten, N., Chatwani, N., & Uraimov, O. (2020). Detecting deepfake videos: An analysis of three techniques. *arXiv*. https://doi.org/10.48550/arXiv.2007.08517

Rosner, H. (2021, July 17). The ethics of a deepfake Anthony Bourdain voice. *The New Yorker.* https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice

Royle, S. (2021, January 6). ‘Deepfake porn images still give me nightmares’. *BBC News.* https://www.bbc.com/news/technology-55546372

S.2559 - 117th Congress (2021-2022): Deepfake Task Force Act. (2022, May 24). https://www.congress.gov/bill/117th-congress/senate-bill/2559

Savin, J. (2022, October 6). Deepfake porn is on the rise – and everyday women are the target. *Cosmopolitan.* https://www.cosmopolitan.com/uk/reports/a41534567/what-are-deepfakes/

Sayler, K. M. & Harris, L. A. (2022, June 3). Deep Fakes and National Security. *Congressional Research Service*. https://crsreports.congress.gov/product/pdf/IF/IF11333

Serwotka, I. (2022, October 24). Mr Deepfakes can make you a porn star. *UnHerd.* https://unherd.com/2022/10/mr-deepfakes-can-make-you-a-porn-star/

Somers, M. (2020, July 21). Deepfakes, explained. *MIT Sloan*. https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

Stump, S. (2021, December 28). Man behind viral Tom Cruise deepfake videos calls the technology ‘morally neutral’. *Today.* https://www.today.com/news/man-tom-cruise-deepfakes-tiktok-speaks-ethics-technology-rcna10163

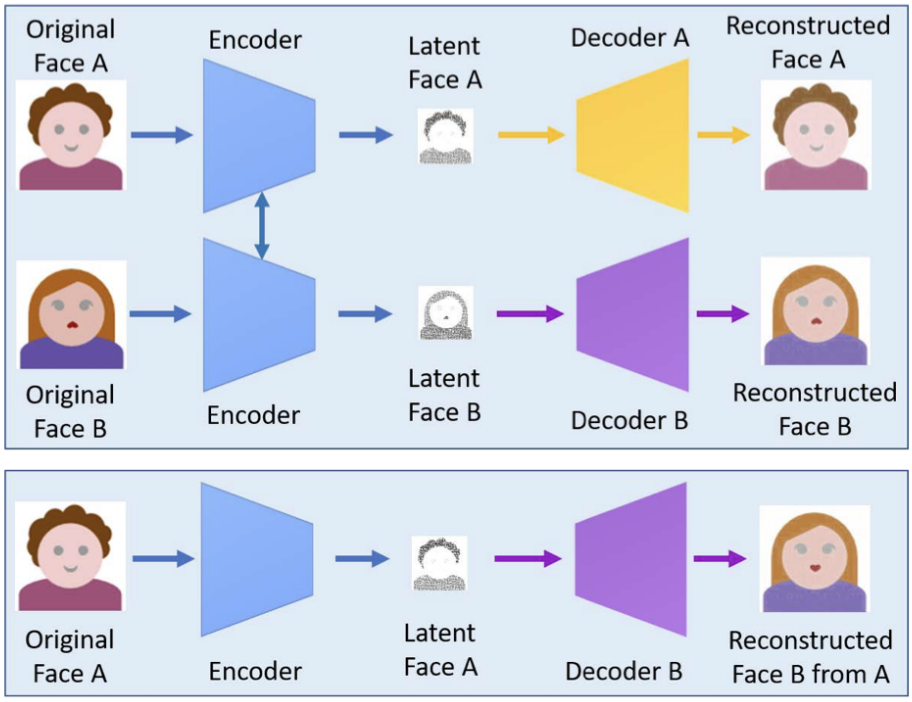
Toews, R. (2022, October 12). Deepfakes are going to wreak havoc on society. we are not prepared. *Forbes*. https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=3f6b61057494

Yang, M. (2021, July 16). Anthony Bourdain documentary sparks backlash for using AI to fake voice. *The Guardian.*

https://www.theguardian.com/food/2021/jul/16/anthony-bourdain-documentary-ai-voiceover-roadrunner

Figure 1

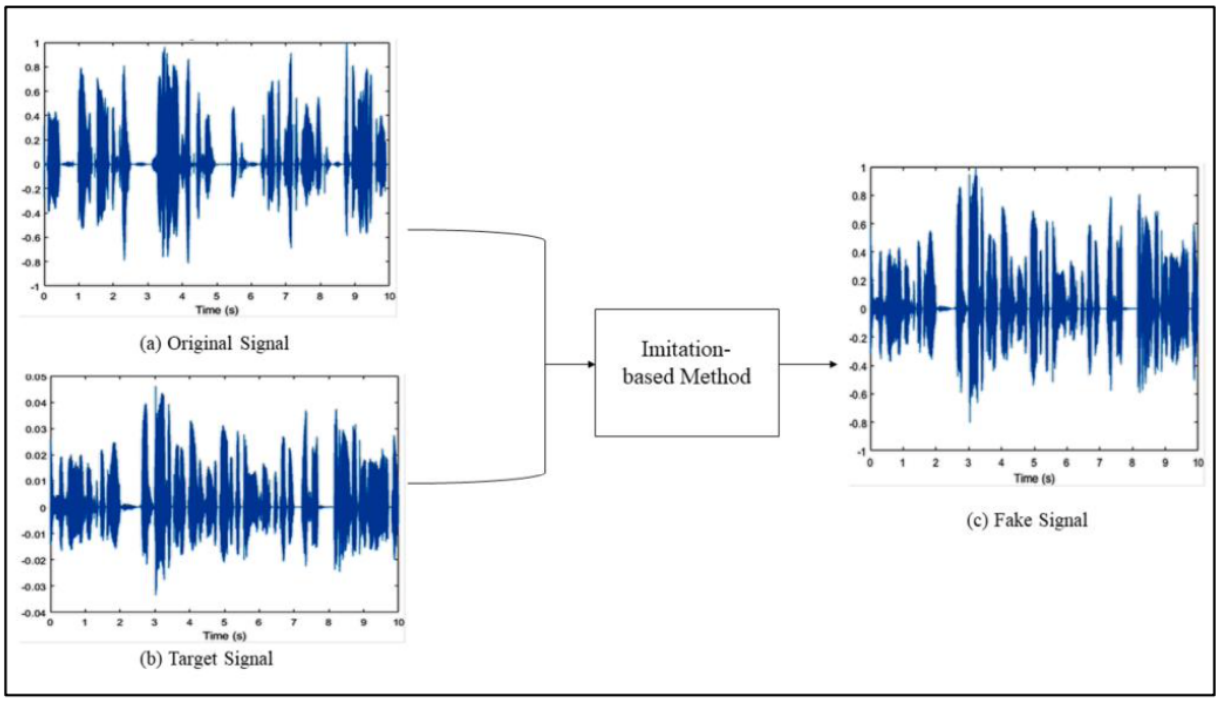
A deepfake creation model using encoder-decoder pairs



Source: Deep Learning for Deepfakes Creation and Detection: A Survey (2022)

Figure 2

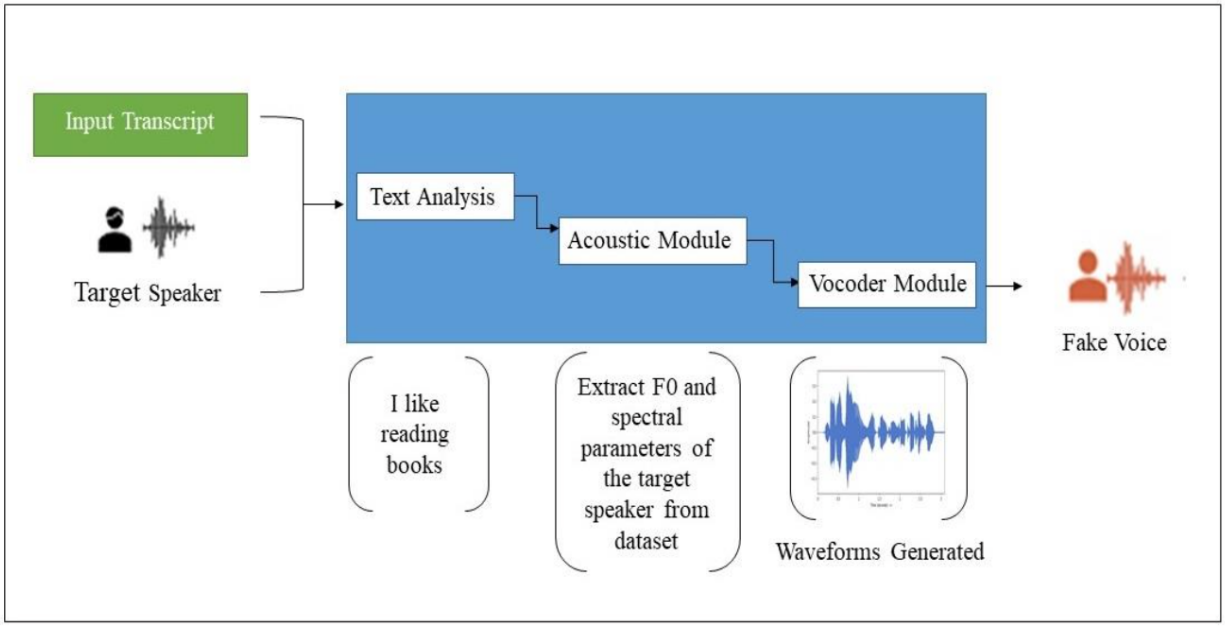
Imitation-based deepfake audio



Source: A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions (2022)

Figure 3

Synthetic-based or text-to-speech deepfake audio



Source: A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions (2022)