Artificial Intelligence Enabled – Deepfake Technology
The Emergence of a New Threat

by

Valencia A. Jones

A Capstone Project Submitted to the Faculty of

Utica College

May 2020

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Cybersecurity

Abstract

The rise of deepfake and disinformation is growing exponentially since its inception three years ago. A combination of *deep learning* and *fake*, deepfakes use machine learning and artificial intelligence technologies to manipulate audio and video media to create fake media content with the high potential to deceive. Cybercriminals are using deepfakes to connect with our emotional biases to alter our way of thinking. The innovation of deepfake technology has been grossly misused to defame an individual's character, create political havoc, disseminate fake news via social media platforms, and threaten stability and democracy globally. There is wide concern that foreign adversaries may seek to threaten and influence the U.S. 2020 electoral process with deepfake disinformation. Disinformation is being used as a weaponization to cause harm and compromise a corporation's reputation. The digital age of deepfakes is growing in sophistication each day and the potential of spreading misinformation via social media platforms is far and wide, eroding the trust citizens have in digital online media content. While there is a great concern about the problematics of deepfakes, the technology has the potential to be used for the good of humanity. Ultimately, as new technologies are being developed to detect and suppress deepfakes, cybercriminals will continue to develop more advances and new forms of threats to maintain a step ahead. To ensure development to protect the nefarious use of deepfake technology, governments, businesses, and academia must work together to create the proper infrastructure for our information ecosystems to grow.

Keywords: Cybersecurity, Professor Duane Corbo, national security, authenticity, cyber, warfare.

Acknowledgments

I would like to acknowledge both Professor Duane Corbo and Dr. Leslie Corbo for your invaluable guidance that you both have provided throughout this research. I would also like to thank the professors that continued to encourage me through this unfamiliar program. Your professional guidance, encouragement, and unwavering support are very much appreciated. Finally, I would like to express my profound gratitude to my son Kenny and daughter Laticia who have always been my greatest supporters in all my accomplishments. As a single parent, you both supported and encouraged me through the years in my educational endeavors and continue as I accomplish achieving another master's degree. A special thank you to my son who encouraged me to open the door to the world of cybersecurity. As you prepare to retire from the military and continue your education, remember, when one door closes, another will open. It is never too late to continue learning.

Table of Contents

List of Illustrative Materials

Statement of the Problem

Deepfake technology has taken on a new level of sophistication, allowing for the unauthorized insertion of audio and visual images into social media platforms, making the differentiation of *real* and *fake* information more challenging than ever before (Greengard, 2020). Deepfake technology not only poses individualized threats to citizens but also erodes trust in our institutions, as well as having national security implications (Westerlund, 2020). Deepfake continues gaining momentum in information warfare as individuals are choosing to believe what they want rather than what they hear and see on social media (Dack, 2019).

Deepfake videos have evolved rapidly through the social media platform in the political sphere to disgrace politicians (Fischbach, 2020); however, the biggest manipulated content has been primarily porn which has been accessible over the Internet for years (R. Metz, 2019; Porup, 2019). According to a study conducted in December 2018 by Amsterdam-based deep learning company Deeptrace Labs, 96% of uploaded videos were pornographic in nature and featured primarily women. According to the study, approximately 14,678 deepfake videos are circulating on the Internet. Deepfake videos continue to grow significantly in both the public and private sectors (R. Metz, 2019). With the presidential election looming upon us in 2020, researchers and lawmakers are concerned about the spreading of disinformation (C. Metz, 2019a). Government officials and politicians are equally concerned that deepfake videos will be used as misinformation to sway citizens, disrupting the political landscape (R. Metz, 2019). The impact of deepfake videos could provide cybercriminals a new means to illicit cyber activity with the spread of disinformation among businesses through sophisticated phishing attacks (“Fake News,” 2020).

The purpose of this research was to examine the implications of artificial intelligence (AI) generated deepfake audio and video technology on digital cyber ecosystems. Specific questions addressed in this research include: What are the primary threats posed by deepfake technology? What are the potential benefits of deepfake technology? How are deepfakes detected and what countermeasures can be implemented in preventing attacks?

Digital technology has historically been available and used in the movie industry to alter images, videos, and audio (J. Kietzmann, Lee, McCarthy & T. Kietzmann, 2019). Deepfake technology has become a growing trend in the new era of digital impersonation increasingly becoming realistic and to the human eye, very convincing (Citron & Chesney, 2019). Deepfakes go beyond the expensive Adobe Photoshop software used to alter an image which is becoming a thing of the past with the next generation of artificial intelligence (AI) synthetically creating fake images that look completely real (C. Metz, 2019b). Individuals are using the AI machine learning platform to manipulate images and videos representing 'fake news' to influence public opinion, blackmail, and video fake porn (Greengard, 2020). Deepfake content also has the potential to impact criminal and personal injury cases by falsifying key evidence (Greengard, 2020).

In the age of digital technology, we are faced with a paradigm shift in digital representation with the advancement in machine learning. Social media platforms are helping in spreading fake media with the speed and ease of sharing content to citizens (Guo, Yasan, Yao, Yunji, & Yu, 2019). Research shows growth in online content viewing, as many individuals are taking to social media platforms to view online news content. In a survey of 5,000 U.S. adults conducted July 8-21, 2019 by Pew Research, 55% get their news from social media platforms either *often* or *sometimes* (Shearer & Grieco, 2019). The survey showed that 52% of Americans

use Facebook to get their news. In addition, respondents stated that one-sided news and inaccurate news is a *very big problem* when it comes to getting news on social media. Even as many Americans are turning to social media platforms for news, many are showing skepticism of 'fake news' content (Westerlund, 2019). It is becoming more difficult to distinguish between authenticated content and fake videos leaving a negative impact among citizens (Greengard, 2020).

The rise of deepfake technology has become a major cybersecurity threat not only to our society, but also to many businesses, our constitutional political system, and national security (Westerlund, 2019). Journalists and news outlets are finding it difficult to distinguish deliberate disinformation news from real news which is diminishing media trust among citizens (Westerlund, 2019). Individuals are choosing to believe what they want rather than what they hear and see on social media (Dack, 2019). The spreading of misleading or manipulative political propaganda used to influence personal views is a major threat to our democracy and our electoral process (Westerlund, 2019).

Overview of Deepfake Landscape

A form of artificial intelligence, the term deepfake is a combination of the term 'deep learning' and 'fake' (Littell, 2019). The deepfake phenomenon originally emerged in late 2017 on the Reddit community discussion website targeting celebrities when a user of the same name posted a digitally altered pornographic video clip of a well-known celebrity with that of a porn actress that essentially sparked an increase in fake videos in media outlets (Littell, 2019). Since its inception, deepfake technology has been used to manipulate videos and voices superimposing face images from an original authentic source onto an existing AI-generated image or video, matching expressions and tone with the intent to create a fake video that appears authentic

(Owen-Jackson, n.d.). One of the more famous deepfake videos was created by actor and comedian Jordan Peele that depicted former President Barack Obama delivering a public service announcement regarding the threat of deepfake technology in an effort to raise awareness of fake news. The video was intended to be humorous, but also to raise awareness of the growth of deepfake technology (Owen-Jackson, n.d.).

The style-based Generative Adversarial Network (GAN) deep learning architecture is modeled on the method of image manipulation, producing fake content, primarily face frontalization (Yin, Jiang, Robinson, & Fu, 2020). The core of deepfake technology is dependent on two main principles of machine learning algorithms. The two competing models of artificial neural networks, 'generator' and 'discriminator' work together creating a GAN (Westerlund, 2019). Artificial neural networks mimic the human brain in learning and recognizing patterns with the concept of the more real image samples that are fed into a dataset, the more accurate it can be replicated as a fake sample. The generator produces the synthetic images creating fake videos from the original dataset (real samples). The discriminator attempts to analyze and distinguish the deepfake video from the real to synthesized for authenticity (Dack, 2019). The cycle continues with the two machine learning algorithms with the generator network continuing to create fake videos until the discriminator network no longer can detect and authenticate the video forgery (Bisen, 2019).

With the proliferation of deepfake technology and the plethora of easy to use sophisticated open-source applications on the market, individuals can manipulate existing videos or images by superimposing someone's face, mimic facial expressions, and synthesize their speech (Westerlund, 2019). What once took weeks and sometimes days and significant skills to

create a convincing deepfake, now only takes a matter of hours to create with widely accessible online applications and software programs (Sample, 2020).

The Age of Hybrid Warfare

Many adversaries and foreign nations have conducted disinformation campaigns for years to deceive people and divide and weaken foreign affairs. In 2014, U.S. residents were targeted by the Russian Internet Research Agency (IRA) in a campaign of disinformation to deceive citizens and popular news media (Padgett, 2019). The IRA had employed individuals to acquire social media accounts of citizens in an attempt to turn U.S. citizens against the American government with fake and deceptive online content posts. The campaign continued through the Presidential elections in 2016, with the IRA crafting their campaign of disinformation with fabricated social media content. In today's digital era, Russia continues to promote hybrid warfare operations as a standard practice using automated artificial intelligence to generate synthetic media to support their disinformation campaigns (Padgett, 2019).

Deepfakes potentially pose a threat to elections, democracy, and international affairs with threat actors circulating fake videos of politicians making statements and doing things that never happened to taint the public's perception of the individual during the election polls (Paris, 2019). Military leaders are concerned that deepfake videos pose a national security threat among forces with the potential of doctored images posted on media outlets to discredit the U.S. government with hostile information. This was a case in 2018 during a military training exercise in the Baltics when American soldiers were claimed to have killed a local child in a collision on a roadway (Rempfer, 2018). There was a roadway collision among four U.S. Army Stryker vehicles when the lead vehicle braked too hard to avoid an obstacle in the road ahead. A doctored image showed a mangled bicycle lying next to a child's deceased body and

unconcerned U.S. soldiers nearby. A blog post made it appear that a popular Lithuania news outlet was claiming the Americans had killed the child instead, all to discredit the training exercise and rally against allies' defense support (Rempfer, 2018).

As deepfake technology progresses, our government's national security may be destabilized, affecting the forthcoming elections. Not only are political candidates at risk of being targeted by deepfakes, but their campaign workers may be subject to bribery for information. Once a political rumor is released, the greater it could impact the outcome of the elections (Fischbach, 2020).

Deepfake Technology and the Business Sector

Cybercriminals have been targeting businesses through social engineering attacks for years using forms of spam, malware, and launching various phishing campaigns to manipulate individuals in divulging confidential information and exploiting weaknesses (Ajder, 2019; Fruhlinger, 2020). Deepfake technology is just another threat avenue making detection more difficult. The proliferation of deepfakes is progressively evolving as new developments in technology emerge and cybercriminals become more sophisticated in developing new and innovative attack vectors (Eddy, 2019).

Social engineering attacks are on the rise and have been linked to fraud in the financial industry with threat actors using voice phishing as a form of deepfake audio deception for extortion (Miles, 2019). Cybercriminals have used artificial intelligence software to manipulate and mimic voice deepfake. In the first known fraud case in March 2019, attackers used AI to impersonate the voice of a chief executive officer (CEO) of a German parent company to scam $243,000. The CEO of a U.K. based energy firm thought he had been on the phone with the parent company CEO when he was asked to make a transfer of funds to one of their suppliers;

however, he became suspicious when funds had not appeared and also when prompted by the third call for another transfer (Livni, 2019). The use of deceptive voice audio to mimic the parent company CEO's voice had been taken from various platforms to include YouTube videos, TED talks, and other recordings and inserted into the AI program to create an audio deepfake (Soare, 2019).

Deepfake technology can potentially cause damage to the stock market with the manipulation of stock prices (Soare, 2019). As with video deepfake and politicians, cybercriminals are using artificial intelligence to alter video statements of influential people such as CEOs and other VIPs of large known companies. These misleading modified media content or sham when distributed, may cause a panic in the market with either a company's stock price to plummet or rise based solely on the fabricated content and reputation (Soare, 2019).

With the advancement of technology and the use of social media, many individuals are oblivious to the fact that when they upload images and videos to the cloud and other online platforms such as Facebook and Instagram, they are placing themselves in a vulnerable state of potentially being victimized by deepfakes (Stout, 2019). Facebook accounts for over 300 million photos uploaded daily, with eight billion videos being viewed. In less than 10 years of existence, over 40 billion photos have been shared on Instagram with an average of 95 million daily uploads (Stout, 2019).

Defining the Audience

As the growing popularity of AI-generated deepfake technology increases with the availability of open-source applications, the audience of this research paper may benefit from understanding the role social media platforms pose with the spreading of disinformation and methods to combat deepfake through education awareness and training. Deepfakes will not

dissipate but will continue as a critical challenge in weaponizing information. The research paper may also help individuals recognize the tell-tale signs of deepfakes, emphasizing that not everything on the social media platform is real, and the importance of questioning and verifying, when possible, the authenticity of online social media content. This research will also serve as a knowledge tool when considering uploading personal images and videos to online cloud-based and social media platforms such as Facebook and Instagram, and the potential of becoming a target to deepfake face-swapping.

Literature Review

Deepfake technology is rapidly proliferating in manipulating audio and video to cause harm and abuse to our society, political system, and businesses (Westerlund, 2020). With the unauthorized insertion of audio and visual images into social media platforms, it has become difficult to differentiate *real* and *fake* information, making it more challenging than ever before (Greengard, 2020). Deepfake technology not only poses individualized threats to citizens but also erodes trust in our institutions, as well as having national security implications (Westerlund, 2020).

The research for the literature review will discuss the primary threats posed by the use of deepfake technology in our digital cyber ecosystems. The research will discuss how it directly affects democracy and national security, private sectors, and social media. The second segment will explore the potential benefits derived from the use of deepfake technology particularly in the educational, entertainment, and healthcare industries. The last segment will conclude with information on how to detect deepfakes and what countermeasures can be implemented in preventing attacks.

Threats Posed by Deepfakes

The weaponization of cyber-enabled information has become an existential threat to political, cultural, and privacy norms questioning our basic idea of reality, crippling the validity of trustworthiness and credibility of information in our ecosystem (Lin, 2019). Deepfakes take on a different form of targeted hacking – a treacherous attack on our brain. Threat actors use this type of manipulation to connect with our emotional biases with the intent to alter our way of thinking and what we believe in. (McIntyre, 2019).

A threat to democracy and security. As deepfake technology evolves, our government's national security will be at the forefront of hybrid warfare of weaponized

disinformation primarily used to destabilize and undermine our forthcoming elections (Fischbach, 2020; Westerlund, 2019). Since the inception of deepfake technology, a plethora of videos has been manipulated to mimic facial expressions and synthesize speech to make it appear as if an individual may be saying or doing things that never actually happened. One major threat to our democracy that is a concern to the intelligence community is the spreading of misleading or manipulative political propaganda used to influence personal views in our electoral process (Westerlund, 2019). According to the 2019 Worldwide Threat Assessment of the US Intelligence Community, U.S. foreign adversaries such as Russian, China, and Iran will almost certainly use deepfakes to manipulate images, audio, and video content in an attempt to influence the outcome of the 2020 presidential election campaign (Coats, 2019).

Deepfakes have the potential of discrediting political candidates and opponents' reputations with the increased use of social networks, provoking a huge political conflict with manipulated content of controversial or disqualifying remarks and messages leading to public discourse (Fischbach 2020). When misinformation is propagated, it tends to spread faster and wider in social networks so once distributed, it is nearly impossible to suppress the impact in the outcome of the elections (Fischbach, 2020). One of the most prominent deepfake videos that gained widespread attention in April 2018 was created by Buzzfeed CEO Jonah Peretti and actor/comedian Joran Peele depicting former President Barack Obama insulting President Donald Trump. The video clip was intended to deliver a public service announcement regarding the threat of deepfake technology and how easy it is to manipulate video content to spread misinformation (Owens-Jackson, n.d.; Shao, 2020). A snapshot image of the digitally manipulated video is shown in Figure 1 below depicting former President Obama on the left and actor Jordan Peele on the right.

Figure 1. Video image of Obama voiced by Peele in PSA message. Reprinted from "*A viral video that appeared to show Obama calling Trump a 'dips---' shows a disturbing new trend called 'deepfakes'"* by K. Fagan, 2018. Retrieved from https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4

Figure 2 shows an image taken from the video created by Peele of former President Obama demonstrating elements of facial mapping.

Figure 2. Image extracted from the video featuring President Obama showing elements of facial mapping. Reprinted from "*The future of the deepfake – and what it means for fact-checkers"* by T. Hwang, 2018. Retrieved from https://www.poynter.org/fact-checking/2018/the-future-of-the-deepfake-and-what-it-means-for-fact-checkers/

Impact on the military. Deepfakes continue to be an instrument of wartime deception manufacturing claims of hostile information by using manipulated evidence to make the claims of civilian casualties more plausible during military operations. In the interim, these claims could prompt enemy recruitment and retaliation as well as discredit ally forces and exercises (Citron &

Chesney, 2019). U.S. military leaders are concerned that deepfake videos will pose a threat to national security and military forces with fabricated images circulating media platforms in an attempt to discredit the U.S. government with hostile information (Rempfer, 2018). In 2018, during a training exercise in the Baltics, a roadway collision took place among U.S. military vehicles to avoid an obstacle in the roadway; however, a doctored image surfaced on a blog post of a Lithuanian news media outlet claiming U.S. soldiers had killed a local child in the collision. It was a deliberate attempt by the Russian leadership to condemn the participation and strengthening of U.S. military and joint allied forces in the region (Rempfer, 2018).

Deepfakes have the potential to threaten the stability and democracy internationally well as leading to nation-state conflicts. In 2018, on the west coast of Central Africa, a video of Gabon's president, Ali Bongo had spread throughout the country setting off mass confusion as to the realism of the video and the president's competence as a leader (Cahlan, 2020). The president customary gives a New Year's address to the country; however, with his recent health issues and stroke, he had not appeared in public (Hao, 2019). His well-being caused the citizens to question his ability to lead the country. To calm the speculation, Gabon's government aired a video of the president giving the customary address raising suspicion among citizens of the validity of the video and whether it had been manipulated. Consequently, the video content led to their military attempting to launch a rebellion against the government a few days later to undermine authority (Joplin, 2019; Panic, 2019).

A growing threat to the private sector. Deepfake video and audio content are becoming a growing cybersecurity threat for many businesses and society. Threat actors are finding new ways to perpetrate fraud attacks through social engineering campaigns (Ajder, 2019). Threat actors are using voice phishing as a form of deepfake audio deception. Through artificial

intelligence (AI), voice manipulation is becoming more popular and damaging than video deepfake since there is no visual content involved for an individual to determine if the call is a scam or not (Gustavsson, 2019). Artificial intelligence disinformation campaigns threaten to damage the reputation of targeted companies that deal with a variety of investors, customers, and potential employees ("Moody's says," 2019).

Extortion. With the technology of video deepfake, cybercriminals are using manipulated voices for impersonation to extort money from employees. Employees have a hard time determining if an audio call has been tampered with since they cannot see the individual on the end, making it harder to determine if the caller they hear is 'real' or 'fake'. (Gustavsson, 2019). In the first known voice fraud case reported by the Wall Street Journal in March 2019, fraudsters used AI voice manipulation to impersonate a parent company's CEO to scam $243,000 in a transfer of funds to one of their suppliers. It wasn't until the third call for another transfer before any suspicion was indicated that it was a voice phishing scam (Livni, 2019). With the current state of the country brought on by the recent COVID-19 crisis, many financial businesses are forced to work remotely in a more relaxed environment. This may prompt threat actors to take advantage of voice phishing techniques to impersonate a CEO to scam a corporate finance officer (CFO) or financial colleague into making transfers ("Deepfake attack," 2020).

Fraudsters are using video conference platforms such as Skype to use synthetic impersonation to extort money and obtain information (Adjer, 2019). From 2015-2017, Israeli con men impersonated the French Foreign Minister over Skype to scam targets for an estimated $90M stating it was for secret operations and to pay ransom for the release of hostages held in the Middle East. The con men created a replica of the foreign minister's office and hired makeup experts to disguise them to successfully extort money (Adjer, 2019; Schofield, 2019).

Fraud. As deepfake technology spreads in the digital realm, there could potentially be a breach of sensitive personal information. Spoofing attacks will be used to compromise face recognition-based biometrics used to authenticate a user's identification. Although convenient for many users for authentication, face biometrics can be spoofed using deepfake digital manipulated videos to face swap – superimpose an individual's face onto an existing video to fraudulently access their account during authentication process (Wojewida, 2020).

Stock market manipulation. The dissemination of disinformation has plagued our economy affecting stock market prices. Deepfakes can potentially cause damage to the stock market with the manipulation of stock prices. Cybercriminals are using artificial intelligence to alter video statements of influential people such as CEOs and other VIPs of large known companies (Soare, 2019). One example was when a false *tweet* had spread through the social media platform claiming former President Barack Obama had been injured in an explosion, resulting in $130 billion in stock value being wiped out. (Vosoughi, Roy, & Aral, 2020). In 2019, Tesla CEO Elon Musk became a victim of deepfake market manipulation when a video of him spread through social media portraying erratic behavior and smoking marijuana (Adjer, 2019). Manipulated media content when distributed, may cause a panic in the market with either a company's stock price to plummet or rise based on reputation (Soare, 2019). Consequently, Tesla's stock dropped 6% (Adjer, 2019).

Social Media and Disinformation. Disinformation and the stigma of fake news social media platforms have become the norm for the consumption of information. At the same time, they can be a menace in spreading reproduced misinformation at a substantial pace impacting those who are targets of deepfake (Albahar & Almalki, 2019). With the advancements in web-based applications and machine learning AI, deepfakes can be created and spread across social

media platforms making it more difficult to reduce potential threats by cybercriminals ("Moody's says," 2019). There is ongoing skepticism among citizens of the dissemination of online content that is suspected of being manipulated from real content to fake to sway their trust in information other than from what they view on their social network. Citizens tend to view content as reliable when received by family members, other relatives, and close friends that share the same view without confirmation of it being manipulated to fake content (Westerlund, 2019).

Many hobbyists view AI-generated images or videos as a form of humor, entertaining, or politically satirical rather than a threat to individuals. In the age of digital technology, open-source software is readily available to individuals to create deepfake content whether they have the technical know-how or expertise in manipulating videos with face swapping of original content or use synthesized audio to change the speech of an individual (Westerlund, 2019). Deepfake videos are produced by collecting still images of the target individual's head and then use open-source software to produce life-like videos (Gertz, 2018).

According to the Reuters Institute Digital News Report, the use of Facebook news has deteriorated since 2016 in many countries to include a 20-point decline among younger Americans (Kuper, 2018). Individuals in developing countries use Facebook's WhatsApp to receive news from friends they associate with; however, most news is false, but friends have trust in what others say. Kuper (2018) indicated that with the use of artificial intelligence and manipulated videos to create deepfake video content, individuals will no longer be able to identify real or fake content or trust in what they see or hear with a percentage of them not caring either way.

Real vs fake. In 2019, a manipulated video of House Speaker Nancy Pelosi had gone viral across social media appearing as though she was slurring her words during a news

conference. According to researchers, the video had been slowed down to about 75% of its original speed to give the impression that she was inebriated (Harwell, 2019). Since the tools used were to alter the speed of the video and poorly done, the video was not considered to be a form of deepfake but referred to as a *cheapfake* (Nelson, Simek, & Maschke, 2020).

Another doctored video that sparked a lot of controversy after going viral on social media was that of 2018 Marjory Stoneman Douglas High School shooting survivor Emma González in Parkland, Florida. Emma González became a civil rights advocate and speaker for the #NeverAgain movement at the March for Our Lives rally in Washington, DC campaigning for stricter gun control and school safety in Florida (Mezzofiore, 2018). Also, she and other teen activists were part of an online video and article she wrote for *Teen Vogue's* gun control issue where Emma was shown ripping up a bullseye shooting target (Citron & Chesney, 2019; Gonzalez, 2018). Unfortunately, her message became a smear campaign receiving personal attacks from conservative figures and far-right activists (Arias, 2019; Mezzofiore, 2018). A doctored image was lifted from her original article and manipulated to show her tearing up The U.S. Constitution instead, making her look like she was a threat to civilians Second Amendment rights to bear arms, destroying the foundation as outlined in the Constitution (Arias, 2019; Horton, 2018). The image went viral on the social medium platform receiving 1,500 retweets and 2,900 likes on Twitter within the first few hours of being posted (Mezzofiore, 2018). The original image and doctored image are shown in Figure 3 below.

Figure 3. Side by side view of the original image of Emma González on left and doctored image on the right. Reprinted from *"No, Emma Gonzalez did not tear up a photo of the Constitution"* by G. Mezzofiore, 2018. Retrieved from https://www.cnn.com/2018/03/26/us/emma-gonzalez-photo-doctored-trnd/index.html

Misinformation and *Fake News*. The term fake news has been used as a universal term in politics as a division between true and false. It has become a declamatory weapon in the political sphere perceived as power and dominance, a means of reinforcing authority and self-preservation (Farkas, 2020). The proliferation of fake news became a global phenomenon during the 2016 presidential election race between Hillary Clinton and Donald Trump when Trump claimed they were false accusations made against him, made up of fake news (Pringle, 2020). It has become the focal point on social media and across the Internet for propagating both misinformation and disinformation to mislead the general public, making it difficult to determine real from fake news. Pringle (2020) defines fake news as stories that are false, deliberately fabricated, but not 100% accurate with no verifiable facts. Fake news can also be used as *clickbait* to intentionally mislead the reader into clicking on an advertisement which the writer would profit from (Pringle, 2020). Vosoughi, Roy, & Aral (2020) believe the term has been irredeemably divided with conflicting information in the current political and media environment and our understanding of what constitutes the idea of news – fake news, false news, rumors, half-truth, etc.

Our information ecosystem is full of false information circulating on social media platforms such as Facebook and Twitter. Misinformation is false or inaccurate information created and shared inadvertently with no intent to deceive. Disinformation is false information deliberately created and shared with the intent to obscure the truth. The COVID-19 pandemic is receiving a plethora amount of false information to the point it is being labeled as a dangerous *infodemic*. This causes social disorder and division among citizens and hinders the response from public health (Vanderslott, 2020). Images and videos continue to be manipulated and disseminated across social media. Figure 4 shows an image of seven types of fake content being shared and the intent (Wardle, 2017).

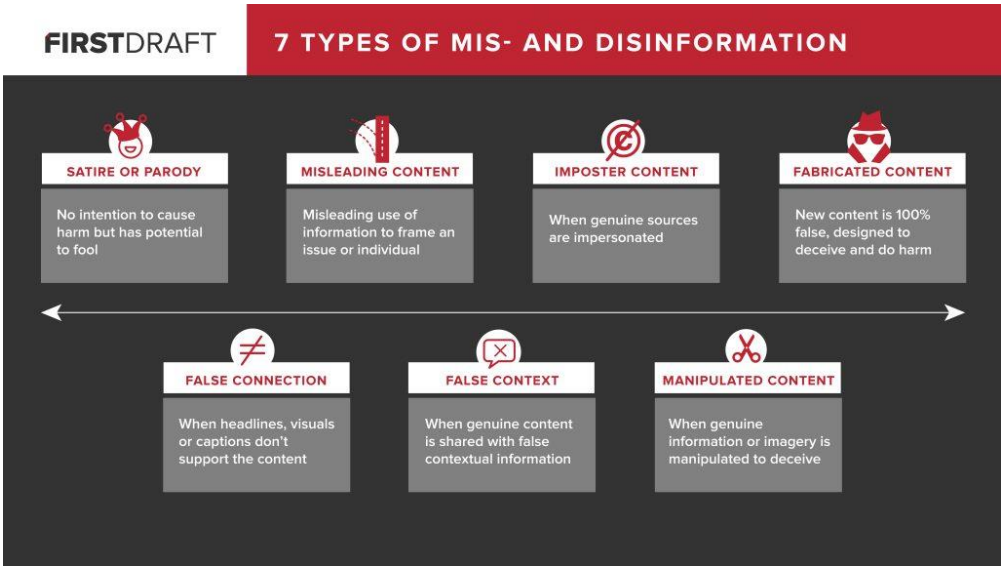

Figure 4. Classification types of fake content. Reprinted from *"Fake news. It's complicated."* Retrieved from https://firstdraftnews.org/latest/fake-news-complicated/

Benefits of Deepfakes

Although most deepfakes have been identified as having a negative impact on our society causing a variety of havoc, it can also be a positive influence in our society ("Seeing is," 2019). There has been so much attention drawn on the consequences of deepfake technology that we lose sight of how AI-based technology can actually be beneficial (Rees, 2019). With the new era

of technology and AI, deepfake technology can produce many beneficial practices in several industries to include education, entertainment, and healthcare (Westerlund, 2019).

Education. The use of deepfakes in the education realm as part of their classroom teaching is a benefit of replicating facial and vocal features. Old lecture recordings of individuals no longer living have been used to emphasize teaching by using the expertise of historical figures in their classroom settings through AI ("Seeing is," 2019). Deepfake technology was used in an exhibition at The Dalí (Salvador Museum) in St. Petersburg, Florida. Using machine learning, the exhibit featured a life-size of The Dalí delivering a variety of quotes he had spoken during interviews throughout his career (Chandler, 2020)

Entertainment. With deepfake technology, movie makers can recreate classic scenes or create new movies using actors that are no longer living. Deepfakes can also be used to enhance techniques used in modeling and simulation, as well as training and education (Miller, 2019). In a 2019 global malaria awareness campaign, the industry used deepfake technology to create a video to attract diverse audiences in different languages. The ad featured David Beckham delivering the video in multi-languages created with visual and voice-altering technology (Westerlund, 2019).

Healthcare. Deepfake technology raises new possibilities in the social and healthcare field. With the advances in artificial intelligence (AI), deepfake technology can help individuals diagnosed with Alzheimer's disease to potentially remember their past with the interaction video contact of them with a younger facial appearance (Westerlund, 2019). AI deepfakes can be used to recreate images through video and audio manipulation of style and speech of recent lost loved ones for individuals to reconnect with them to aid in their grieving process.

With the advancement in medical technology and the use of deepfake technologies,

medical professionals are better enabled to develop digital tools such as recreating an amputee's limb or the capability of a transgender individual to view themselves as a different gender. Scientists are using the benefits of medical diagnosis by exploring generative models to detect medical abnormalities in x-ray images (Westerlund, 2019). In 2019, over 41 million patient records were breached with patient privacy increasing from 450 security incidents in 2016 to 572 incidents in 2019 (Landi, 2020). With the increase in data breaches and patient privacy, advancements in AI-based technology in the healthcare field could enable researchers to create an entire imaginary population of patients in a single hospital to develop and test a variety of ways to diagnose, monitor, or treat diseases with lifelike data without jeopardizing patient privacy (Rees, 2019).

Detection and Countermeasures

In this era of information warfare and the sophistication of deepfake technology, it is becoming increasingly problematic to verify the authentication of video content. The Department of Defense (DoD) is positioned with the burden of finding solutions to counter it to protect our national security (Strout, 2019). With the current elections coming up and proliferation of deepfakes, researchers from academic institutions, tech firms, and nonprofits are vigorously working on developing tools to detect misleading AI-generated deepfake video content. The deepfake arms race is just at the beginning stages of a short-term solution (Wiggers, 2020).

How to detect deepfakes. It has become very difficult, if not nearly impossible, to distinguish between a real video and a fake video alone with the human eye (Bisen, 2019). There are still flaws in the deep learning platform in detecting manipulated content; however, researchers have determined that there are some significant ways to detect AI-generated fake video content (Chawla, 2019). To manipulate content, there has to be a good amount of video

content available on the Internet for machine learning to take place (T. Nguyen, C. Nguyen, Dung Nguyen, Duc Nguyen, & Nahavandi, 2019). Since deepfake algorithms typically do not create video content with normal blinking patterns – blinking is a lot less, it is easier to detect a 'fake' from 'real'. Most humans blink anywhere between every 2 to 10 seconds, with a single blink taking between one-tenth and four-tenths of a second with very few images found with an individual's eyes closed (Chawla, 2019; Nguyen et al., 2019). Other methods include unnatural movements of the body between the face and head, a noticeable change of lighting in the background, and the skin tone of the video image considerably smooth (Bisen, 2019). Using AI-based technology in video magnification techniques could determine if an individual in a video was actual or computer-generated by detecting changes in the human pulse. (Susarla, 2018).

Digital photos have become an integral part of communications. With the popularity of social media platforms such as Facebook and Twitter, manipulated video and audio content proliferates becoming a big issue of trust in what is fake and what is real. Individuals can detect and verify the source of deepfakes before they share news on social media with available open-source or free image editing and manipulation tools for analyzing photoshopped images. Anyone with a computer or smartphone can use a set of manipulation tools to detect a deepfake. Additionally, fact-checking websites such as *AP Fact Check* and *Full Fact* identify some of the most recent common fake news stories (Vanderslott, 2020). Table 1 below lists four image detector software programs, the platform used to access, and a brief description of capabilities (Tengyuen, 2020).

Table 1

Image Detector Software Programs Used to Detect Deepfakes

Software	Options	Platform	Key Features
FotoForensics	Free, Paid	Web	Uses an advanced algorithm to decode any possible photoshopped pictures and manipulation; it uses the Error Level Analysis (ELA) to identify areas within an image that are at different compression levels.
JPEGSnoop	Free, OpenSource	Windows	Examines and decodes the inner details of JPEG, MotionJPEG AVI, and Photoshop files. It can also be used to analyze the source of an image to test its authenticity.
Ghiro	OpenSource	Linux	Digital forensics analysis tool that extracts information from images in a report; fully automated.
Forensically	Free	Web	Set of free tools for digital image forensics. It includes clone detection, error level analysis, metadata extraction, and more.

Methods used to detect deepfakes. Deepfakes have increasingly become harmful to our democracy and security, and society (Nguyen et al., 2019). Most methods proposed to detect deepfakes are based on deep learning and artificial intelligence. Researchers continue to explore potential technological solutions to detect and aid in the negative consequences of deepfakes (Dack, 2019).

Convolutional neural networks (CNNs) use a machine learning algorithm to detect deepfake videos. Once performed by humans, the age of artificial intelligence (AI) takes images and videos and uploads them into systems for the machines to learn facial expressions and determine whether uploaded content is authentic of ‘fake’. The algorithm runs in the background, monitoring the system for any synthetic content uploaded then either removing the manipulated content or alerting the user before being shared on social media platforms (Albahar & Almalki 2019).

Watermarking is based on a neural network that discretely embeds artifacts directly in a video or audio file without distorting the contents. These artifacts are sensitive to tampering and

will identify hidden traces of manipulated content. To ensure integrity verification when shared on social media platforms, these artifacts would alert the recipient of the fake content (Albahar & Almalki, 2019). For this method to work, devices used for capturing the video and audio content must be capable of creating a digital watermark. In addition, social media platforms must authorize the posting of these digital watermarks on their sites (Dack, 2019).

Defense measures to combat deepfakes. Preventive measures to deepfakes will require the development and deployment of anti-deepfake technology. The technology includes a combination of deepfake detection, content authentication, and deepfake prevention. Other methods include legislative and regulation to address civil or criminal laws; education and training to raise public; and establishing corporate policies (Westerlund, 2019).

Anti-deepfake technology. The Intelligence Advanced Research Project Activity (IARPA) has been sponsoring research programs to develop technologies to detect facial biometric technologies (Kimery, 2018). At the guidance of Congress and the recent defense policy bill signed in December 2019, IARPA received a $5 million prize security challenge to invest in community researchers to develop defensive measures that will detect deepfakes and validate their authenticity automatically (Keller, 2020). The DoD has initiated the Defense Advanced Research Projects Agency (DARPA) Media Forensics (MediFor) program to accelerate the development of automated AI-based technologies that will assist in detecting and assessing the integrity of videos and images to determine if they have been manipulated (Strout, 2019). Google, as well as other technology companies, are stepping up to assist researchers in developing tools to combat deepfakes (Rees, 2019).

Legislative and regulatory measures. Lawmakers face challenges in national security and democratic governance. Enacting laws will threaten First Amendment protection. To combat

the efforts of distribution channels for deepfakes through social media platforms and video hosting sites, the Deepfake Report Act of 2019, if passed by the House and then President, will require the Department of Homeland Security (DHS), the Department of Justice, Federal Election Commission, the National Science Foundation, and others, to address deepfakes and their impact on national security on an annual basis (Miller, 2019; Ruiz 2020).

Education and training measures. Businesses should be aware of the latest use of audio content to scam individuals into authorizing wire payments of company money to pay for goods and services as many employees are oblivious to audio deepfakes. Providing a platform for educating and training employees against the defenses of such attacks could be beneficial in motivating an individual to question the validity of an unusual request for payment or network access (Nelson et al., 2020). Also, companies should consider monitoring their social media platforms and websites for deepfake content, educate employees, and incorporate the latest technology to combat and protect their business ("Moody's say," 2019).

Individuals need to be aware of the possibility of becoming a victim to various forms of communication, an important fact in verifying authenticity before performing any action that may require monetary action through voice mimicry. In addition to current two-factor authentication (2FA), individuals that receive a voicemail message to perform an action should verify the validity of the requests. This may require the individual to make a dial-back to the number from the incoming call to verify (Lorica & Loukides, 2019).

Other methods to combat deepfakes. Government agencies should invest in tools specifically designed to detect deepfake by looking at the data content of the displayed changes that typically are not seen by the naked eye. (Fischbach, 2020). Not all employees are expected to recognize a deepfake; however, web and security solutions can prevent interactions and

unusual activity at the first onset with extra checks at the business process level. Fischbach (2020) says that using basic behavior monitoring of an employee can be essential when they have fallen for a deepfake. Many of their actions will deviate from normal baseline analyses that can be tracked and flagged, locking them from their account (Fischbach, 2020).

Digital lenders are combatting the impact of fraud attacks and taking countermeasures to spot deepfake videos. As part of a loan qualification method, photos are taken from the property site during inspections and surveys for individuals applying for loans online. The lenders are using high-quality software programs to verify the authenticity of the photos submitted (Crosman, 2020).

Discussion of the Findings

The purpose of this research was to examine the implications of artificial intelligence (AI) generated deepfake audio and video technology on digital cyber ecosystems. Specific questions addressed in this research include: What are the primary threats posed by deepfake technology? What are the potential benefits of deepfake technology? How are deepfakes detected and what countermeasures can be implemented in preventing attacks?

Primary Threats Posed by Deepfake Technology

A political weapon. Threat actors have been using a myriad of videos to manipulate content, mimic facial expressions, and synthesize audio to create fake videos that appear authentic (Westerlund, 2019). Accordingly, deepfakes potentially pose a threat to elections when fake videos circulate through media platforms, changing the reality of politicians making statements to discredit them and taint the public's perception during election polls (Fischbach, 2020; Paris, 2019). The intelligence community has expressed a similar concern since the 2016 presidential campaign, warning about the threat of foreign adversaries meddling in the 2020 election process (Westerlund, 2019). Fischbach (2020) concluded that the threat of deepfakes will have an impact on our national security as it continues to grow and improve, stating that the implications could threaten our political sphere in the upcoming 2020 electoral process. However, according to a 2019 Report from Amsterdam-based Deeptrace Labs, no identified instances of deepfakes used in disinformation campaigns were found to have been used for political recourse leading to the 2020 U.S. presidential race; nevertheless, it was determined that the technology could be used to create deepfakes (Hao, 2019).

Adversaries such as China, Russia, Iran, and North Korea have continuously posed a threat to U.S. national security interest with threats of cyber espionage, critical infrastructure

disruption, along with Russia's interference to influence the 2016 presidential election (Coats, 2019; Padgett, 2019). In 2019, Former Director of National Intelligence Dan Coats provided an overview of the national security threats posed to the U.S in the Worldwide Threat Assessment of the US Intelligence Community report. In his statement to the intelligence community, he indicated that deepfake technology could pose a major threat to the U.S. in 2020 by altering the information we rely on upon through social media. The report further indicated that U.S. foreign adversaries will likely use deepfake to manipulate images, audio, and video content to interfere, weaken, and disrupt the 2020 presidential election infrastructure and outcome (Coats, 2019).

The 2018 manipulated video clip of former President Obama spoken by actor/comedian Jordan Peele provided a prime example of the threat advanced deepfake technology poses on national security. The synthesized audio and video showed the former President sharing his concern about U.S. adversaries and the ongoing threat to U.S. national security using deepfake technology by saying, "We're entering an era in which our enemies can make it look like anyone is saying anything at any point in time." (Owens-Jackson, n.d.; Shao, 2020).

The country of Gabon has been riddled with a democratic discourse with the foundation built on a family legacy of a 43-year long presidency. With his absence from the public for months before his customary New Year's address, the citizens suspected that the government was covering up his state of health, questioning the country's political future (Hao, 2019). Bongo and his government were more concerned with protecting the family legacy instead of addressing his health issues to displace the rumors and suspicions of his competency to lead. As a result, the controversial 2018 video of Gabon's President Ali Bongo delivering his address threatened the stability and democracy of the country. The video displayed an unsettling look of the president to include his eyes rarely moving or blinking, one perceived determination of a deepfake video by

researchers. Consequently, his video address questioned by citizens and military as being manipulated, launched an unsuccessful rebellion against the government (Hao, 2019; Joplin, 2019). According to research, University of Albany professor of computer science and digital-media forensics expert Siwei Lyu ran a forensics analysis on the video clip through his deepfake algorithm, concluding that the contents had 99% probability of authenticity (Cahlan, 2020).

Citron and Cheney (2019) argued that deepfakes are a threat to national security that could inadvertently have systemic dimensions. A 2019 doctored video of House Speaker Nancy Pelosi went viral on Facebook depicting her as fumbling over her words during a news conference. The video had been viewed more than two million times and shared more than 45,000 times within the first 24 hours (Harwell, 2019). Originally considered to be a deepfake; however, research showed that the video had been debunked as using AI to alter the content. Instead, the video was considered to be a *cheapfake* since the timing of the video was altered by slowing down the speed (Nelson et al., 2020). Whether considered a cheapfake or deepfake, future altered videos could influence public perception and gravely impact our society.

Threats posed to private sectors. Research indicates much discussion and concern has been about the dangers deepfake technology and disinformation poses on our national security and societal platform; however, disinformation and digitally manipulated media content can pose a serious threat to our businesses, financial markets, and corporate reputations as well. More commonly are threats of market manipulation, extortion, fraud, and social engineering attacks (Ajder, 2019). Market manipulation attacks may incite panic in the market with the dissemination of disinformation affecting stock market prices thus impacting our economy. Westerlund (2019) indicated that deepfakes could influence the financial market by announcing false mergers, spreading video content of a CEO making racist or other prejudicial slurs, or the

making of a false statement of financial losses or bankruptcy. Social engineering attacks to include extortion and fraud, have been prevalent as a threat to businesses for years with attacks on a company's IT infrastructure; however, with the emergence of deepfakes, cybercriminals are using a more sophisticated approach to cyberattacks by using AI-generated and machine learning algorithms to mimic reality to supercharge these attacks (Ajder, 2019).

Research shows that voice manipulation is becoming more popular and damaging than video deepfake, primarily because they typically have no visual content involved making it difficult to determine authenticity. Voice phishing or 'vishing' is becoming a new methodology of deception in deepfake attacks and often used to enhance business email compromise attacks by impersonating individuals to fraudulently extort money from businesses (Livni, 2019; Miles, 2019). Miles (2019) indicates that voice manipulation of source audio can be created in under four seconds with all the distinguishable traits used from a multitude of audio clips of a target, thus fed into the combined AI and machine learning algorithm to create a convincing deepfake audio.

Research indicated that the use of deepfake audio as a deceptive method has been found in at least three successful attacks whereby a company CEO's voice had been mimicked to extort money transfers estimating in the millions of dollars. In a March 2019 report by the Wall Street Journal, a fraudster used AI-generated audio to impersonate the voice of a CEO of a German-based parent company to illicitly scam $243,000 from a U.K. based energy company requesting the funds be transferred to one of their suppliers (Levni, 2019). It was through three calls from the fraudster, one to initiate the transfer, a second to state the funds had been reimbursed, and the third call to request another transfer, before the CEO of the U.K. company became suspicious of the scam. The money was transferred to an account in Mexico before the CEO had realized he

had been a victim of a scam when the reimbursed funds had not appeared in the account (Damiani, 2019; Gustavsson, 2019).

Current research revealed that even amid the COVID-19 crisis, fraudsters are targeting small businesses that are financially affected with the pandemic by offering fake loan programs. As with the attack on the U.K. based energy company, threat actors may use deepfake voicemails to scam executives by impersonating business leaders in making a transfer of funds or deceiving employees into taking nefarious actions especially in a time of crisis and their need to help ("Deepfake attack," 2020). Since many CEO's have social media profiles with audio content, it would be relatively easy for a threat actor to obtain voice samples of the victim to impersonate and scam an individual ("Deepfake attack," 2020).

The battleground of social media and fake news. Social media platforms such as Facebook and Twitter have become the norm for the consumption of news and information for many individuals but could be a platform for the spreading of disinformation. Online news and posts continue to surface through social media platforms by individuals that reshare information, sometimes manipulating the content into fake news. Unfortunately, our information ecosystem is saturated with false information continuously exploited by threat actors with the means to intensify the spread of disinformation making it more difficult to discern real from fake. According to Farkas (2020), fake news has become a declamatory weapon in the political landscape causing widespread debate and an attack on opponents. The only way society will be able to combat fake news is through verifying the validity of the contents since many individuals receive and share content they believe is from a trusted source.

The concept of fake news has been in existence since the early 13th century with the spread of made-up news and propaganda about the Battle of Kadesh in Egypt. It wasn't until the

rise of information warfare, that the term, ‘fake news’ evolved with the 2016 presidential election race between Hillary Clinton and Donald Trump. Since then, the usage of the term has escalated to the reckless disregard of the truth making it difficult to decipher real from fake news. The electoral landscape became a global phenomenon when Trump claimed false accusations of him that spread across the Internet was made up of fake news that was deliberately fabricated to defame his character (Pringle, 2020).

Fake news can contain both misinformation and disinformation; however, the two do not share the same meaning. Pringle (2020) and research identify both as false information; however, misinformation is the notion of creating and sharing false information inadvertently with no intent to cause harm, while disinformation is the intent to obscure the truth. Research illustrates that fake news is spread faster and further than real news, with fake news stories reaching 1,500 views, six times faster than real stories, and is only getting worse. According to Michigan State University information systems professor Anjana Susarla, countries with relatively high levels of economic growth could essentially encounter more fake news than real information by 2022 (Susarla, 2018).

Recent research shows that fake news concerning the COVID-19 pandemic is spreading faster than the virus itself to an extent that it is causing social disorder and division among citizens (Vanderslott, 2020). Dack (2019) pointed out that deepfakes will increase in influencing our view of fake news while decreasing our immediate connection to facts that are shared causing a great deal of confusion. Research continues to show a shared perception indicating that in the era of deepfakes, if individuals can’t trust what they see and hear online themselves, then they will ultimately choose what they want to believe, or not caring either way.

Gab network that provides a platform for free speech, had posted a doctored image lifted

from an original video of 2018 Parkland Florida high school shooting survivor Emma González that went viral on social media when the image posted perceived that she was ripping up the U.S. Constitution. Her original image was taken from an article in *Teen Vogue's* issue on gun control showing her ripping up a bullseye shooting target (Mezzofiore, 2018). To illustrate how quickly false information propagates, several research articles showed that the image had received 2,900 likes and 1,500 retweets on Twitter within the first few hours of being posted on the social media platform. Gab intended to present the image as a parody/satire to discredit the students who lead the #NeverAgain movement against gun violence. Consequently, Gab received criticism from Twitter users for publishing the fake news and NRA supporters received backlash in a tweet from University of Wisconsin, professor Donald Moynihan who debunked the fake image, criticizing supporters for displaying the surviving teenagers at a threat to citizen's rights to bear arms (Arias, 2019). Cintro and Chesney (2019) indicate that the growth of deepfake video and audio content will present a threat to society and create significant policy and legal challenges.

Potential Benefits of Deepfake Technology

With technology continuously evolving, there is always going to be risks like all new technology, but the flip side of the coin, there can be new innovative ideas and positive implications that can help society. There has been much emphasis placed on the dark side of deepfakes and their negative impact on our society, that we tend to overlook the beneficial facets the technology may have in envisioning future possibilities of AI technology. With the advancement of AI and machine learning, deepfake technology may serve to be advantageous to many industries enabling new forms of communication.

According to research, deepfakes can and have been used for educational purposes in both the classroom and the arts for audience engagement. Several pieces of research showed that deepfakes can help the public engage in the education of history and culture through the arts with

the life-size realistic deepfake exhibit of Salvador Dalí (The Dalí) delivering a variety of quotes he had spoken or written during his career as an artist. Chandler (2020) added that it took 1,000 hours of machine learning to create the deepfake audio content of Dalí's old interviews and stories about his life. Citron and Chesney (2019) shared that educators have the opportunity to provide students with educational innovation with historical figures speaking directly to students during a lecture. Classroom lectures can be supplemented with old lecture recordings of historical figures such as Albert Einstein using both facial and voice features with his old recordings to enhance a better understanding of his theory instruction ("Seeing is," 2019). Research indicated that the flipped classroom model of outside classroom teaching may be part of the future in deepfake to deliver instructional videos in multiple languages.

Another emerging use of deepfake technology in education and entertainment is using commercial products in multiple languages to attract a diverse audience around the world. The U.K. software company Sythesia used synthetic media to create an anti-malaria awareness campaign video featuring David Beckham speaking in nine different languages (Westerlund, 2019). Using AI and voice-altering technology, the entrepreneurs created a 3D model of Beckham, re-animated to appear as if his voice was used in all languages (Barr, 2019).

According to research, deepfakes have been used by movie makers to recreate classic scenes or create new movies or sequels using actors that no longer are living. Deepfakes can also be used to create movies with current actors using their characters from previous film footage presented into new or updated scenes. Also, deepfakes have been used in a variety of movies using automatic and realistic voice dubbing in different languages to deliver to a diverse audience (Westerlund, 2019).

The benefits of deepfake methods also extend into healthcare research raising new possibilities. Rees (2019) explains that AI deepfakes could enable researchers to create an imaginary population of virtual patients and data to improve patient care without using actual patient data. Instituting AI-generated deepfake technology will become an important tool in leveraging patient care in the healthcare industry to develop in assisting with recognizing diseases, medical diagnoses, treatment techniques, and patient outcomes. Other benefits include the possibility of helping individuals with diseases to connect and interact with others in the same state to improve their quality of life. Scientists are exploring the possibility of advance medical science and medical discoveries by creating virtual chemical molecules using deepfake technology (Westerlund, 2019). The future possibilities in AI-generated deepfake could have the potential to grow exponentially in improving patient care.

Detecting and Implementing Prevention Measures of Deepfakes

Detecting AI-generated deepfake content. Detection is an important suppression tool and although a deepfake video may appear to be authentic, manipulated content can be detected. Research shows that there is a growing interest in detection tools based on the potential impact that deepfakes may pose. Detecting manipulated video and audio deepfakes can often present a challenge as researchers continue to explore potential technological solutions; however, research has shown there are different methods in determining what is real and what is fake video content. Common research methods to detect face manipulation are missing details in visual features - eye blinking patterns, eye color, and reflections in teeth. Research indicates that even with AI-generated deepfakes, little details such as eye blinking are difficult to manipulate. Nguyen et al. (2019) indicate that the rate of eye blinking in a deepfake is less frequent than the normal 2 to 10 seconds. Researchers further shared that many of the images found online rarely displayed

individuals with their eyes closed, which makes it difficult for the deepfake algorithm to generate a fake video capturing normal blinking; however, a high rate of blinking may also indicate tampered video content. Other face manipulations include detecting the differences of visual artifacts in eye color and missing details in teeth area. Media forensics experts suggest other methods of detecting deepfakes could include reflections and shadows in the background (Westerlund, 2019).

Most methods to detect deepfakes are based on AI and deep learning machine learning techniques using different types of neural networks. A team of researchers from the New York University School of Engineering developed an AI-watermarking technique based on a neural network to embed artifacts or create immutable metadata onto a video or audio file permitting easy recognition and verification when manipulation of contents has taken place (Albahar and Almalki, 2019; Guedim, 2019). The researchers tested this approach identifying the chances of deepfake detection of manipulation to be between 45 to over 90 percent without distorting the contents (Guedim, 2019). Research shows that CNNs make face detection more stable and face alignment more reliable by learning facial expressions of uploaded content to systems. The algorithm will detect fake content and either remove it or alert users before dissemination on social media platforms (Albahar and Almalki, 2019).

With the proliferation of video deepfakes spreading across social media platforms such as Facebook and Twitter, it has become difficult to ascertain which content to trust as a legitimate news source. The only way that society can combat fake news is verifying the validity of the contents since many individuals receive and share content they believe is from a trusted source; therefore, individuals can be proactive in propagating misinformation through social media platforms. Since many individuals have either a computer or smartphone, they can use readily

available open-source software or free manipulation tools from online repositories without any technical knowledge to detect a deepfake before sharing through their social media. Additionally, research showed a variety of fact-checking websites such as *AP Fact Check* and *Full Fact* are available for checking and verifying the accuracy of information. These fact-checking sites identify some of the most recent fake news stories disseminated across the Internet and social media platforms (Vanderslott, 2020).

Implementing countermeasures. With continued advancement in AI and machine learning, threat actors and adversaries could potentially use deepfake technology to create highly advanced threats to cripple our economy, society, and security. As deepfake technology increases in sophistication, methods will need to be implemented to combat any threats to our information ecosystem. To defend against deepfake threats, methods will need to be implemented in anti-deepfake technology; legislation and regulation; corporate policies; and education and training (Westerlund, 2019).

In the growing age of deepfake technology, the Defense Department is pushing for the government and other industries to develop anti-deepfake technology tools that would detect, authenticate, and prevent deepfakes before they become a major national security issue. Research has shown that in 2018, the DARPA produced the first forensics tools in AI to tackle deepfakes. DARPA's MediFor program was created to develop a comprehensive set of tools to assess whether videos or images had been manipulated and completed, as well as address both current and developing capabilities (Strout, 2019). Further research showed that DARPA has already spent an estimated $68 million on digital forensics since 2016. As the arms race continues, the $5 million IARPA received based on the 2020 Defense Authorization Act will help the Defense Department in their challenge to researchers in academia and technology industries to develop

AI tools to find ways to counter the threat of deepfakes. Google recently released 3,000 deepfake videos to assist researchers in developing methods to identify and combat deepfakes (Rees, 2019). However, as AI develops and deepfakes get better, the countermeasures will have to grow as well to keep pace. According to Hany Farid, a synthetic content expert at the University of California Berkeley, more time, skill, and risk will be needed to counter deepfakes.

Lawmakers are at the beginning stages of legislating deepfakes indicating their concern with the potential that deepfakes have on our society, especially through major media platforms such as Facebook and Twitter that distribute deepfakes. To combat this, the Deepfake Report Act of 2019, currently pending in the US House of Representatives, would mandate the DHS to publish a report annually for five years reporting on deepfake technology. The report would address methods to detect and counter deepfakes, and how threat actors and foreign adversaries are using deepfakes to incite harm on our national security. However, the legislative is unable to act against foreign adversaries to stop political interference. Further, regulating deepfakes will threaten First Amendment protection, limiting what lawmakers can do to protect our democracy from manipulations and disinformation campaigns. Research shows that several states have enacted or working to legislate deepfake laws to suppress audio and video disinformation that would make it unlawful to distribute; however, they too are faced with free speech rights (Ruiz, 2020). Citron and Chaney (2019) argue that prohibiting deepfakes entirely would take away the experimentation in the various fields of history, science, art, and education.

As technology grows, it is going to be relevant that individuals, education institutes, and businesses are educated on the implications that deepfakes may pose on society. Proper education, awareness, and training will be essential in combatting deepfakes in the future. For businesses, providing a platform for educating employees about the emerging threat of deepfakes

and incorporate the latest technology will help in preventing damage to their reputation. Amid the COVID-19 pandemic, threat actors may use the opportunity to attempt to spread discord and disinformation and seek financial gain through fraud attacks. Businesses will need to provide awareness of the implications of audio deepfake threats to their finance employees encouraging them to use the *four eyes* principle check - two individuals will review and double-check an action before it is approved ("Deepfake attack," 2020). Additionally, businesses should integrate social media and website authentication tools to combat and protect their business interest ("Moody's say," 2019). Westerlund (2019) suggests that education institutes implement a program to teach critical thinking and digital literacy that will help students in the ability to spot fake news. Also, students should be educated on the trustworthiness of online information as the differentiate real from bake. Westerlund (2019) further suggests that social media companies invest in available detection technologies to prevent their platform from being weaponized with disinformation. Additionally, they will need to work together to flag suspected content, block, and remove known deepfakes to help from the content being reposted in a user's news feed (Westerlund, 2019).

Digital lenders have already incorporated countermeasures to detect deepfakes used in fraud attacks. Lenders are intentionally inviting cybercriminals to apply for loans online to catch them while others are taking other security measures such as the deployment of technology that will be able to spot deepfake videos. A 2019 study by LexisNexis Risk Solutions identified an 8.2% increase in fraud attacks on digital lenders from the previous 24 months (Crossman 2020).

Comparative Findings to Other Research

As with previous research findings, the deepfake phenomenon is a major concern to our society. Previous research concentrated on technological issues and the negative impacts that AI-

generated deepfakes pose on society; however, this research takes a look at the current landscape of digital content manipulation and our ecosystem. The research highlights the positive benefits the AI-generated deepfake technology may have on society and humanity in the future within various industries. Additionally, this research discussed the impact deepfake has on social media platforms and how society is willing to accept less than truthful information.

Limitations to Research

Deepfake technology is still in the early stages of development as well as the technologies to detect them. Although a plethora of articles on deepfake technology detection and prevention were reviewed, a limited number of scholarly research articles that addressed or examined the current state of deepfakes were available. Research demonstrated how voice manipulation was damaging; however, there were no clear definitive studies on detection methods and countermeasures. With the technology coming into public awareness only two years ago, there were no clear solutions to define the long-term impacts of deepfake. As the dangers of deepfakes become more known and understood, other future development may change the overall threat level.

Conclusion

The purpose of this research was to examine the implications of artificial intelligence (AI) generated deepfake audio and video technology on digital cyber ecosystems. Specific questions addressed in this research include: What are the primary threats posed by deepfake technology? What are the potential benefits of deepfake technology? How are deepfakes detected and what countermeasures can be implemented in preventing attacks?

Deepfakes are not a new technology, doctored media has been around for years. Only now is the technology gaining momentum as software apps are making it easy for anyone to create deepfakes and use them for nefarious purposes. The makeup of deepfakes combines AI,

machine learning algorithms, and GAN tools that allow users to create manipulated video and audio content. While the generator algorithm is like a *forger* creating fake content, the discriminator algorithm attempts to detect whether the content is real or fake.

Since the onset of deepfake technology, threat actors have been using a myriad of videos to manipulate content to make it appear as if an individual may be saying or doing things that never actually happened. Using content to mimic facial expressions and synthesize audio to create fake videos, cybercriminals are using this type of manipulation to connect with our emotional biases with the intent to alter our way of thinking and what we believe in (McIntyre, 2019). Accordingly, there is a growing concern that manipulated video and audio content will be used to spread misleading political propaganda to influence citizens' personal views in the U.S. 2020 electoral process (Westerlund, 2019). The intelligence community has also expressed a similar concern that deepfakes will have a significant impact on our national security and warn that foreign adversaries such as Russia may be a major threat in influencing the presidential election (Coats, 2019); Westerlund, 2019). However, as reported by Deeptrace Labs in 2019, no identified instances of deepfakes disinformation campaigns have been used to influence the 2020 electoral infrastructure to date.

U.S. military leaders have raised concerns about foreign adversaries using deepfakes as a wartime deception during military operations. Military personnel deployed overseas have been a target of deepfake disinformation conspiracies by adversaries in the past with fabricated images circulating media platforms. Such disinformation could discredit the U.S. government and cause distinction with allied forces in defense strengthening (Rempfer, 2018).

Malicious use of deepfakes has threatened the stability and democracy internationally and in nation-states. One of the known international cases that led to a nation-state conflict with a

purported deepfake, was in the country of Gabon that set off mass confusion of the realism of released video content of their president, Ali Bongo, leaving citizens questioning the authentication that eventually led to an unsuccessful attempt on the government by a military coup (Joplin, 2019; Panic, 2019). Subsequently, forensics analysis never found the video to have been altered or manipulated (Cahlan, 2020; Hao, 2019).

Private sectors have been riddled with fraud attacks and social engineering campaigns for years to target their IT infrastructure. Threat actors are continuing to target business operations with deepfake fraud using AI voice manipulation to extort money from employees. Cybercriminals are using a new methodology to carry out their attacks through voice phishing by impersonating business executives to fraudulently extort money from employees (Livni, 2019). It isn't a difficult feat since many CEO's have social media profiles with audio content that threat actors can use to obtain voice samples to carry out their scam. This deceptive method is becoming more popular and damaging to businesses since there is no visual content involved making it difficult to determine the authenticity of the call (Gustavsson, 2019).

AI-generated deepfake content can pose serious threats to the financial market and affect a corporation's reputation. The dissemination of disinformation could affect the financial market with stock market price manipulation impacting our economy. Cybercriminals have used altered videos of influential people to spread disinformation through social media platforms containing false statements about financial status or making prejudicial or racist comments affecting their reputation and the company's stock price (Adjer, 2019).

Social media platforms have become a major battleground in both misinformation and disinformation with the propagation of fake news to mislead the general public. Both forms of fake news are considered to be false information created and shared across social media;

however, misinformation is shared inadvertently without the intent to cause harm, while disinformation is intended to obscure the truth. Disinformation does not stay stagnant on one platform and can be shared across multiple media platforms at a given time (Pringle, 2020).

The term fake news has been used recklessly in the political arena causing a division between true and false. It became a global phenomenon during the 2016 presidential election race when Donald Trump claimed false accusations of him were spreading across the Internet as made-up or fake news (Pringle, 2020). Since then the term has been used widely making it difficult for citizens to see the difference between reliable information and biased information.

Many consumers tend to use various social media platforms to consume and view news and other information online. In a survey of 5,000 U.S. adults conducted by Pew Research in 2019, 55% of Americans get their news from social media platforms either *often* or *sometimes* with 52% using Facebook to get their news (Shearer & Grieco, 2019). Facebook and Twitter have been used by consumers to spread fake news and then reshared among their friends, family members, and other relatives without any authentication influencing the group of individuals (Westerlund, 2019). Research showed that fake news stories are spreading faster than real news reaching 1,500 views six times faster. This type of media saturation makes it nearly impossible to discern between real and fake.

Our society is full of information consumed and propagated through social media platforms with deepfakes causing damaging effects on those individuals who are targets of video content. The doctored video of high school shooting survivor Emma Gonzalez tearing up the U.S. Constitution was liked 2,900 times on Twitter within the first hour of posting and retweeted 1,500 times inciting a smear campaign against her and a personal attack. A manipulated video of House Speaker Nancy Pelosi showing her stumbling over her words, labeled false at the time,

went viral on social media retweeted by President Trump as a personal attack against her (Harwell, 2019).

Amid the negative impacts that deepfakes pose, there are positive implications to draw upon with future possibilities in the arts and history, education, and healthcare. Research has shown that deepfakes can help the public engage in history and culture through the Dalí Museum's life-size deepfake display of artist Salvador Dali delivering audio content of his quotes from his career and stories about his life. Educators can use AI-generated deepfake technology to provide students with supplemental educational innovation using historical figures such as Albert Einstein to speak directly to students during a lecture (Cintron and Chesney, 2019). According to research, deepfake technology can be used in the film industry to recreate classic scenes or create new movies or sequels using actors that are no longer living. Additionally, visual and voice-altering technology can be used to attract a diverse audience around the world in different languages (Westerlund, 2019).

With the advances in AI, deepfakes can be used for the good of humanity to augment and complement the technology in the healthcare domain, rather than replace it (Rees, 2019). Deepfake technology can help in assisting with recognizing diseases, medical diagnosis, and treatment techniques or explore generative models to detect medical abnormalities in x-ray images (Westerlund, 2019). The future possibilities of AI-generated deepfakes could have the potential to grow exponentially in improving patient care.

As deepfake technology expands and becomes sophisticated, it will be difficult to distinguish between a real video and a fake video alone, eventually becoming indistinguishable to the human eye (Bisen, 2019). Often you can spot a deepfake by looking closely at movements and facial expressions. The DoD is in an arms race to find solutions to protect our national

security. The DoD has initiated IARPA and DARPA's MediFor to accelerate the development of automated AI-based technologies by soliciting researchers from academic institutions, tech firms, and nonprofits to develop tools to detect, authenticate, and prevent deepfakes. This is only a short-term solution and a challenge for detection method development, but as AI develops and deepfakes get better, future research will need to focus on more robust, scalable, and generalizable methods for countermeasures to keep pace (Nguyen et al., 2019). In the meantime, individuals with a computer or smartphone can use open-source image detector software programs to detect and verify the source of deepfakes before sharing through social media platforms. Fact-checking websites such as *AP Fast Check* and *Full Fact* are also available for checking and verifying the accuracy of some of the most recent common fake news stories spread across social media platforms (Vanderslott, 2020).

Education is key to any emerging technology. Government authorities, businesses, and education institutes will need to increase awareness of the threats posed by deepfakes to our society and national security. Students should be educated on the dangers of deepfake, where to find reliable information, and how to recognize deepfakes. Proper education and training will be crucial in combatting detecting and combatting deepfakes. Businesses should implement the *four-eyes* principle check for all business transactions before any action is taken to deflate any potential deepfake fraud attack.

References

Ajder, H. (2019). Social engineering and sabotage: Why deepfakes pose an unprecedented threat to businesses. Retrieved from https://deeptracelabs.com/social-engineering-and-sabotage-why-deepfakes-pose-an-unprecedented-threat-to-businesses/

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

Arias, X. (2019). How fake new affected Parkland survivor Emma González. Retrieved from https://medium.com/@xarias_33513/how-fake-news-affected-parkland-survivor-emma-gonz%C3%A1lez-949cc647f75b

Barr, S. (2019). David Beckham appeals for end to malaria by 'speaking' in nine languages. Retrieved from https://www.independent.co.uk/life-style/health-and-families/david-beckham-malaria-must-die-campaign-disease-nine-languages-a8861246.html

Bisen, V. S. (2019). What is deepfake: know everything about this AI-based Technology. Retrieved from https://www.vsinghbisen.com/technology/what-is-deepfake-know-everything-about-this-ai-based-technology/

Cahlan, S. (2020). How misinformation helped spark an attempted coup in Gabon. Retrieved from https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/

Chandler, S. (2020). Why deepfakes are a net positive for humanity. Retrieved from https://www.forbes.com/sites/simonchandler/2020/03/09/why-deepfakes-are-a-net-positive-for-humanity/#4577116d2f84

Chawla, R. 2019. Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4(6): 4–8. Retrieved from https://www.ijarnd.com/manuscripts/v4i6/V4I6-1143.pdf

Citron, D. K. & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Retrieved from https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=1640&context=faculty_scholarship

Coats, D. R. (2019). Worldwide threat assessment of the US intelligence community. Retrieved from https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf

Crosman, P. (2020, Jan 29). Online lenders confront deepfake threat. *American Banker* Retrieved from https://search.proquest.com/docview/2346835496?accountid=28902

Dack, S. (2019). Deep fakes, fake news, and what comes next. Retrieved from https://jsis.washington.edu/news/deep-fakes-fake-news-and-what-comes-next/#_ftnref39

Damiani, J. (2019). A voice deepfake was used to scam a CEO out of $243,000. Retrieved from https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/#de0d18922416

Deepfake attack threat during COVID-19 (2020). Retrieved from http://www.mynewsdesk.com/nccgroup/news/deepfake-attack-threat-during-covid-19-398391

Eddy, M. (2019). Scammers for phishing with deepfakes. Retrieved from https://www.pcmag.com/opinions/scammers-go-phishing-with-deepfakes

Fake news: What could deepfakes and AI scams mean for cybersecurity? (2020). Retrieved from https://www.orange-business.com/en/magazine/fake-news-what-could-deepfakes-and-ai-scams-mean-cybersecurity

Fagan, K. (2018). A viral video that appeared to show Obama calling Trump a ‘dips---’ shows a disturbing new trend called ‘deepfakes’. Retrieved from https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4

Fargas, J. (2020). A case against the post-truth era: Revisiting Mouffe’s critique of consensus-based democracy. In M. Zimdars & K. McLeod (Eds.), Fake news: Understanding media and misinformation in the digital age [Kindle for IPad]. Retrieved from http://www.amazon.com

Fischbach, N. (2020). A National security threat: it’s time to get proactive against deepfakes. Retrieved from https://www.hstoday.us/subject-matter-areas/infrastructure-security/a-national-security-threat-its-time-to-get-proactive-against-deepfakes/

Fruhlinger, J. (2020). What is phishing? How this cyber attack works and how to prevent it. Retrieved from https://www.csoonline.com/article/2117843/what-is-phishing-how-this-cyber-attack-works-and-how-to-prevent-it.html

Gertz, B. (2018). Artificial intelligence to weaponize fake videos. Retrieved from https://freebeacon.com/national-security/artificial-intelligence-weaponize-fake-videos/

Gonzalez, E. (2018). Emma Gonzalez on why this generation needs gun control. Retrieved from https://www.teenvogue.com/story/emma-gonzalez-parkland-gun-control-cover

Greengard, S. (2020). Will deepfakes do deep damage? *Communications of the ACM*, *63*(1), 17–19. doi: 10.1145/3371409

Guedim, Z. (2019). Researchers develop an AI-watermarking technique to spot deepfakes. Retrieved from https://edgy.app/researchers-develop-an-ai-watermarking-technique-to-spot-deepfakes

Guo, B., Yasan, D. Yao, L., Yunji, L., & Yu, Z. (2019). The future of misinformation detection: New perspectives and trends. Retrieved from https://arxiv.org/pdf/1909.03654v1.pdf

Gustavsson, C. (2019). Voice deepfakes: the latest phone scam. Retrieved from tack methodology has remained largely unchanged. Retrieved from https://truecaller.blog/2019/09/23/voice-deepfake-the-latest-phone-scam/

Hao, K. (2019). The biggest threat of deepfakes isn't the deepfakes themselves. Retrieved from https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/

Harwell, D. (2019). Faked Pelosi videos, slowed to make her appear drunk spread across social media. Retrieved from https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/

Horton, A. (2018). A fake photo of Emma Gonzalez went viral on the far right, where Parkland teens are villains. Retrieved from https://www.washingtonpost.com/news/the-intersect/wp/2018/03/25/a-fake-photo-of-emma-gonzalez-went-viral-on-the-far-right-where-parkland-teens-are-villains/

Hwang, T. (2018). The future of the deepfake – and what it means for fact-checkers. Retrieved from https://www.poynter.org/fact-checking/2018/the-future-of-the-deepfake-and-what-it-means-for-fact-checkers/

Joplin, T. (2019). A military coup in Gabon inspired by a potential deepfake video is our political future. Retrieved from https://www.albawaba.com/news/military-coup-gabon-inspired-potential-deepfake-video-our-political-future-1284760

Keller, J. (2020). Researchers eye $5 million program for new technologies in detecting deepfakes. *Military & Aerospace Electronics*, *31*(2), 6–7.

http://search.ebscohost.com.ezproxy.utica.edu/login.aspx?direct=true&AuthType=ip,cookie,url,uid&db=a9h&AN=141870934&site=ehost-live

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135-146. doi.org/10.1016/j.bushor.2019.11.006

Kimery, A. (2018). Deep fake technology outpacing security countermeasures. Retrieved from https://www.biometricupdate.com/201812/deep-fake-technology-outpacing-security-countermeasures

Kuper, S. (2018). The age of skepticism: From distrust to 'deepfake'. *FT.Com,* Retrieved from https://search.proquest.com/docview/2122099097?accountid=28902

Landi, H. (2020). Number of patient records breached nearly triples in 2019. Retrieved from https://www.fiercehealthcare.com/tech/number-patient-records-breached-2019-almost-tripled-from-2018-as-healthcare-faces-new-threats

Lin, H. (2019) The existential threat from cyber-enabled information warfare, *Bulletin of the Atomic Scientists*, *75*(4), 187-196. doi.org/10.1080/00963402.2019.1629574

Littell, J. (2019). Don't believe your eyes (or ears): the weaponization of artificial intelligence, machine learning, and deepfakes. Retrieved from https://warontherocks.com/2019/10/dont-believe-your-eyes-or-ears-the-weaponization-of-artificial-intelligence-machine-learning-and-deepfakes/

Livni, E. (2019). A new kind of cybercrime uses AI and your voice against you. Retrieved from https://qz.com/1699819/a-new-kind-of-cybercrime-uses-ai-and-your-voice-against-you/

Lorica, B. & Loukides, M. (2019). A world of deepfakes. Retrieved from https://www.oreilly.com/radar/a-world-of-deepfakes/

McIntyre, J. (2019). ‘Deepfakes” a national security threat. Retrieved from https://www.washingtonexaminer.com/policy/defense-national-security/deepfakes-a-national-security-threat

Metz, C. (2019a). Internet companies prepare to fight the ‘deepfake’ future. Retrieved https://www.nytimes.com/2019/11/24/technology/tech-companies-deepfakes.html

Metz, C. (2019b). Spot the Deepfake. (It's Getting Harder.). *New York Times*, p. B1(L). Retrieved from https://link-gale-com.ezproxy.utica.edu/apps/doc/A606687291/ITOF?u=nysl_ce_uticacol&sid=ITOF&xid=d9474896

Metz, R. (2019). The number of deepfake videos online is spiking. Most are porn. Retrieved from https://www.cnn.com/2019/10/07/tech/deepfake-videos-increase/index.html

Mezzofiore, G. (2018). No, Emma Gonzalez did not tear up a photo of the Constitution. Retrieved from https://www.cnn.com/2018/03/26/us/emma-gonzalez-photo-doctored-trnd/index.html

Miles, J. (2019). Why deepfakes are revolutionizing the world of phishing. Retrieved from https://www.mimecast.com/blog/2019/10/deepfakes-revolutionizing-phishing/

Miller, J. (2019). The realities and challenges of legislating deepfakes. *Signal, 74*(1), 8. Retrieved from https://search.proquest.com/docview/2287335771?accountid=28902

Moody's says deepfake disinformation campaigns pose reputational risks to businesses. (2019, August 3). *Economic Times*. Retrieved from https://link-gale-com.ezproxy.utica.edu/apps/doc/A595291958/ITOF?u=nysl_ce_uticacol&sid=ITOF&xid=e8bf256f

Nelson, S. D., Simek, J. W., & Maschke, M. (2020). Detecting Deepfakes. *Law Practice: The Business of Practicing Law*, *46*(1), 42–47.

Owen-Jackson, C. (n.d.). What does the rise of deepfakes mean for the future of cybersecurity? Retrieved from https://www.kaspersky.com/blog/secure-futures-magazine/deepfakes-2019/28954/

Padgett, S. (n.d.). The art of digital deception: getting left of bang on deepfakes. Retrieved from https://smallwarsjournal.com/jrnl/art/art-digital-deception-getting-left-bang-deepfakes

Panic, B. (2019). Malicious use of deepfakes is a threat to democracy everywhere. Retrieved from https://medium.com/swlh/malicious-use-of-deepfakes-is-a-threat-to-democracy-everywhere-51a020bd81e

Paris, B. (2019). The deeper danger of deepfakes; worry less about politicians and more about powerless people. Retrieved from https://www.nydailynews.com/opinion/ny-oped-the-real-deepfake-danger-20190920-mnu6w7xbdzgklgfr7ibmq6bcne-story.html

Porup, J. M. (2019). How and why deepfake videos work – and what is at risk. Retrieved from https://www.csoonline.com/article/3293002/deepfake-videos-how-and-why-they-work.html

Pringle, M. (2020). 13 facts about fake news that you need to know in 2020. Retrieved from https://www.wizcase.com/blog/facts-about-fake-news/?keyword=&campaignID=1391397730&matchtype=b&adgroupID=83838117070&adpos=&extension=&kwd=dsa-19959388920&location=&geo=9031304&matchtype=b&device=&ad=406470028999&placement=&adposition=&keyword=&campaignID=1391397730&matchtype=b&adgroupID=83838117070&adpos=&extension=&kwd=dsa-19959388920&location=&geo=9031304&matchtype=b&device=&ad=406470028999&p

lacement=&adposition=&gclid=CjwKCAjwguzzBRBiEiwAgU0FT7pU2mrSVyVg4WU8Kb6sPB5au7CUkuzTxX5BS-53e0cdzwhcnH51qxoCWokQAvD_BwE

Rees, G. (2019). Here's how deepfake technology can actually be a good thing. Retrieved from https://www.weforum.org/agenda/2019/11/advantages-of-artificial-intelligence/

Rempfer, K. (2018). Ever head of 'deep fake' technology? The phony audio and video tech could be used to blackmail US troops. Retrieved from https://www.militarytimes.com/news/your-air-force/2018/07/19/ever-heard-of-deep-fake-technology-the-phony-audio-and-video-tech-could-be-used-to-blackmail-us-troops/

Ruiz, D. (2020). Deepfakes laws and proposals flood US. Retrieved from https://blog.malwarebytes.com/artificial-intelligence/2020/01/deepfakes-laws-and-proposals-flood-us/

Sample, I. (2020). What are deepfakes – and how can you spot them? Retrieved from https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them

Schofield, H. (2019). The fake French minister in a silicone mask who stole millions. Retrieved from https://www.bbc.com/news/world-europe-48510027

Seeing is no longer believing. (2019, Feb 28). *University Wire* Retrieved from https://search.proquest.com/docview/2186816059?accountid=28902/

Seven types of fake news identified to help detect misinformation (2019). Retrieved from https://www.dailyexcelsior.com/seven-types-of-fake-news-identified-to-help-detect-misinformation/

Shao, G. (2020). Fake videos could be the next big problem in the 2020 elections. Retrieved from https://www.cnbc.com/2019/10/15/deepfakes-could-be-problem-for-the-2020-election.html

Shearer, E. & Grieco, E. (2019). Americans are wary of the role social media sites play in delivering the news. Retrieved from https://www.journalism.org/2019/10/02/americans-are-wary-of-the-role-social-media-sites-play-in-delivering-the-news/

Soare, B. (2019). How deepfakes can ruin your business and how you can keep your organization safe. Retrieved from https://heimdalsecurity.com/blog/deepfakes-can-ruin-your-business/

Stout, D. W. (2019). Social Media Statistics 2019: Top Networks by the Numbers. Retrieved from https://dustn.tv/social-media-statistics/

Strout, N. (2019). How the Pentagon is tackling deepfakes as a national security problem. Retrieved from https://www.c4isrnet.com/information-warfare/2019/08/29/how-the-pentagon-is-tackling-deepfakes-as-a-national-security-problem/

Susarla, A. (2018). How artificial intelligence can detect - and create - fake news. *US Fed News Service, Including US State News* Retrieved from https://search.proquest.com/docview/2082561464?accountid=28902

Tengyuen, N. (2020). 4 free fake image detector – analyze photoshopped photos. Retrieved from https://www.geckoandfly.com/10023/analyze-photoshopped-photos-with-fbi-csi-and-cia-fotoforensics-software/

Vanderslott, S. (2020). How to spot coronavirus fake news – an expert guide. Retrieved from https://theconversation.com/how-to-spot-coronavirus-fake-news-an-expert-guide-133843

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151. doi: 10.1126/science.aap9559

Wardle, C. (2018). Fake news. It's complicated. Retrieved from https://firstdraftnews.org/latest/fake-news-complicated/

Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology Innovation Management Review, 9(11), 39-52. Retrieved from https://search.proquest.com/docview/2329154005?accountid=28902

Wiggers, K. (2020). Deepfakes and deep media: A new security battleground. Retrieved from https://venturebeat.com/2020/02/11/deepfake-media-and-detection-methods/

Wojewida, J. (2020). The deepfake threat to face biometrics. *Biometric Technology Today, 2020*(2), 5-7. doi.org/10.1016/S0969-4765(20)30023-0

Yin, Y., Jiang, S., Robinson, J. & Fu, Y. (2020), Dual-attendance GAN for large-pose face frontalization. Retrieved from https://arxiv.org/pdf/2002.07227.pdf