Jack Hepburn

Dr. McDonough

English 3010

7 April 2022

AI and Deepfakes: The spread of misinformation

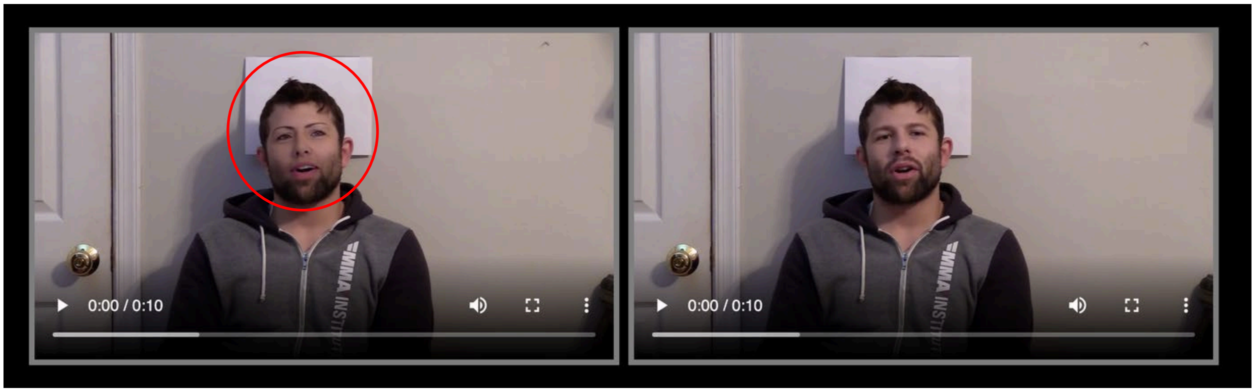
In this day and age, we have the ability to access an abundance of information at our fingertips. Currently, the internet is one of the biggest platforms that we get our news and other information from. While it is very easy and quick to find news and answers, it is oversaturated with news sources and information that might not be credible. “The manipulation of data is not new. Ancient Romans chiseled names and portraits off stone, permanently deleting a person’s identity and history” (Somers). Misinformation and fake news have already been deemed a societal issue but when deepfakes are added to the equation, it is much more difficult to judge the credibility of the source at first glance. The topic of deep fakes is more than just a question of ethics, it is a recent and ever-growing controversy in modern times dealing with the growth of technology and its impact on our society. Being educated on this will prevent us from falling for false media so we must ask the simple questions: How does deepfake technology work and what are the societal implications involved? Are there any ways to counter this technology? Only time will tell.

Deepfakes are considered to be a form of synthetic media and has already been around for years but just recently this technology has emerged into a whole new level of realism. To begin: what are deepfakes? A deepfake refers to deep-learning AI technology that manipulates videos, images, or audio related media. One of the main underlying technologies that enable deepfake creation is a subset of AI known as generative adversarial networks, or GANs (Engler). These algorithms need to be trained with thousands upon thousands of pictures and videos of a person that is the target for the fake. Once the program has compiled enough data from the source media, it can reproduce very convincing and realistic fake videos of an individual on top of another video or image. Because of the power and potential deepfake technology holds, it raises the risk of misinformation and malicious uses. The next question is: Is this technology *really* effective in what it does? Yes and no...

While deepfake technology does have its limitations, there are many scenarios where this technology can be highly effective in. For example, videos that are stationary and have little to no camera movement are susceptible to deepfake manipulation because it is much easier for the algorithm to detect faces when there is less motion. In addition to videos with lack of motion, videos or images that contain less shadows and proper lighting can allow for very realistic deepfakes to be created. Attempting to use deepfake technology on a piece of media where these parameters are not met, will show apparent inconsistencies that give away it is fake. In a publication by Matt Groh of MIT, he dives into the minuscule inaccuracies that can be found in deepfake videos.

> “Pay attention to the cheeks and forehead. Does the skin appear too smooth or too wrinkly? Is the agedness of the skin similar to the agedness of the hair and eyes? Deepfakes are often incongruent on some dimensions. Pay attention to the eyes and eyebrows. Do shadows appear in places you would not expect? Deepfakes often fail to fully represent the natural physics of a scene. Pay attention to the glasses. Is there any glare? Is there too much glare? Does the angle of the glare change when the person moves” (Groh)?

Although there are many more details to lookout for, deepfakes have continually shown themselves to be unable to recreate these natural details of our physical world accurately. "High-end deepfake manipulations are almost always facial transformations" (Groh). Currently, this technology is primarily focused and most effective on facial targets but there is no limit to how these neural networks may evolve throughout history.



(Groh)

Example of a deepfake video compared to the original video created by Matt Groh at MIT. The left side is the fake and the right is the real. Notice the mouth and eyes are not fitting on the left and even includes some extra digital artifacts that take away from the realism.

Currently, it is extremely easy to access and create deepfakes for little to no cost. A program called DeepFaceLab is among one of the most advanced tools for creating deepfakes that is completely free of cost. This program can be utilized to create realistic face swaps over any form of visual media. There are several other tools that can be used to create deepfakes for smartphones as well, but they are not nearly as advanced or as convincing as the tools available for computers. Typically, consumer-grade software for smartphones that are intended for creating synthetic media are not convincing and are meant for entertainment purposes. With computer systems, the software is much more elaborate and can create highly realistic fakes and the only bottleneck is the hardware inside the computer.

In general, deepfakes seem to have a mostly negative stigma behind them. While this is true, there are many ways in which deepfake technology has helped create advances and benefits within education, film industry, art, and accessibility. To begin, this technology has begun to be utilized by VFX artists to help create realistic CGI and bring scenes to life. “AI generated synthetic media can bring unprecedented opportunities in the entertainment industry, and we see a lot of realization of the opportunity by independent creators or YouTube” (Jaiman). For example, in Fast and Furious 7, VFX artists used deepfake technology to recreate a post-humous Paul Walker. The affordability and effectiveness of this technology is a key reason why it has major potential to benefit creative industries. It has even helped independent productions achieve big-budget film quality at a fraction of the price, even free.

Deepfakes utilize artificial intelligence (AI) algorithms that can create a plethora of synthetic media ranging from audio files and pictures to full on videos. These algorithms have recently been utilized to create synthetic voices for people living with ALS. “Imagine being able to talk in your own voice with your loved ones even after losing the ability to speak” (Jaiman). This technology is being worked on by a company called Team Gleason. There are many other uses for this technology for hearing impaired people specifically. In addition, text-to-speech software that is powered by AI and deepfake technology can create unique vocal personas for individuals who have difficulties speaking or are deaf. Many of these technologies are being used to improve daily lives and create accessibility for people with these disabilities. In the education world, these options are creating better interactive and engaging learning environments to those with and without disabilities.

Along with accessibility benefits, deep fake technology is going to start being used in the classroom. Deepfakes can be used to recreate and enhance historical figures or events which can help engage students and create better understanding of events for students. This can be a good thing because it will help us understand the history of the world on a more intimate and relatable level, but it will also come with moral implications. "One obvious danger is the creation of fake historical episodes" (Eisikovits). Creating biased or false outcomes on certain historical events could end up creating even more division and skewed political views within our society.

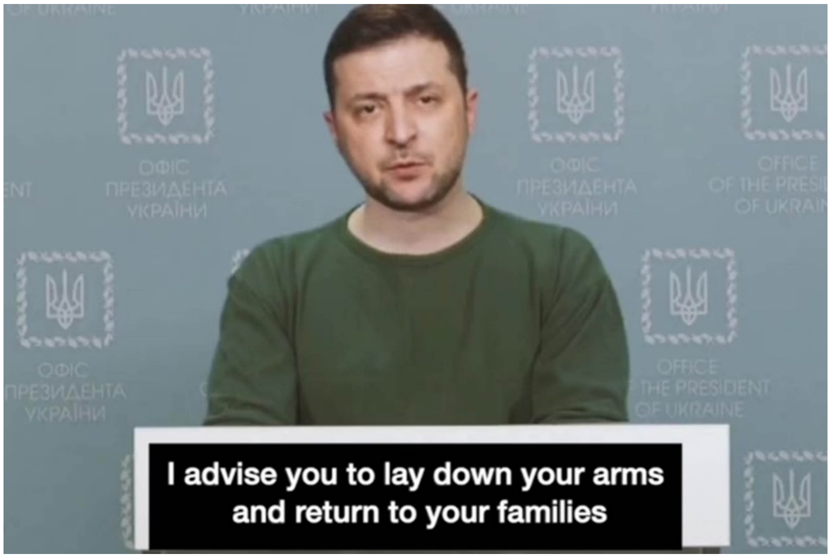
"People have always been better at inventing things than at thinking about what the things they invent do to them" (Eisikovits). While the evolution of AI and deepfakes are revolutionizing our technological potential in many positive ways, there is still great societal damage that can be done if these programs get into the wrong hands. In a study done by Cristian Vaccari and Andrew Chadwick at Loughborough University, they gathered data to see how deceptive deepfakes really were to students on the campus. In total, the study consisted of 2,005 participants in which they were to watch a deceptive deepfake clip of Barack Obama. They were then asked a series of questions asking whether they thought it was real, unsure, or if it was fake.

"Overall, only 50.8% of subjects were not deceived by the deepfake. This finding is surprising given the statement was highly improbable. A smaller, though by no means negligible, group (16%) was deceived, while 33.2% were uncertain" (Vaccari, Chadwick). While the percent of not deceived individuals could have been much worse, half of the individuals that participated were unsure or wrong about the video. Imagine this scenario on a bigger scale; If 50% of the entire population was completely convinced a fake, politically motivated deepfake video was real, there would be a major controversy and divide within society. The study from Loughborough University shows that deepfakes create uncertainty in the media and could be a dangerous weapon to society when in the wrong hands. The Ukraine and Russia conflict is just a glimpse of the harm that can be done by deepfakes.

(Sosa)

Barack Obama deepfake voiced by Jordan Peele example used in this study at Loughborough University.

“Deepfakes pose a significant problem for public knowledge. Their development is not a watershed moment—altered images, audio, and video have pervaded the internet for a long time—but they will significantly contribute to the continued erosion of faith in digital content” (qtd. in Engler). The ease at which these programs can be accessed increase the probability of this technology being used with malicious intent or even weaponized. In recent times, the war between Ukraine and Russia has been one of the most prominent conflicts of within the last 10 years. This conflict has sadly unfolded into an all-out war between the two countries, both sides utilizing weapons and artillery to defend their land. In addition, Russia is believed to have created a deepfake video of the Ukrainian president, Volodymyr Zelenskyy, telling the nation to surrender (Allyn). Hackers posted this fake video of Zelenskyy on a Ukrainian television station and a news broadcast website which could have been devastating if it was not addressed as urgently as it was. Incidents like these show that deepfakes are being weaponized and proves that this technology is already in the wrong hands. This is just the beginning of a much bigger phenomenon that should not be taken lightly.


ОФІС ПРЕЗИДЕНТА УКРАЇНИ
OFFICE OF THE PRESIDENT OF UKRAINE
I advise you to lay down your arms
and return to your families

(Miller)

Deepfake video created by Russia of Ukrainian President, Volodymyr Zelenskyy.

Knowing some of the harmful uses of deepfake technology, how can we limit the damage dealt? There are currently algorithms being developed that are being ‘trained’ to detect deepfakes. “But though strong detection algorithms are emerging (including GAN-based methods), they are lagging behind the innovation found in the creation of deep fakes” (Chesney, Citron). While this technology is using the same method (GAN) that deepfakes utilize, it will never evolve ahead of the deepfake technology. There is potential in this technological innovation to detect deepfakes, but there are several other solutions to this problem that do not include complex algorithms like the GAN methods.

> “Legal and regulatory frameworks could play a role in mitigating the problem, but as with most technology-based solutions they will struggle to have broad effect, especially in the case of international relations. Existing laws already address some of the most malicious fakes; a number of criminal and tort statutes forbid the intentional distribution of false, harmful information. But these laws have limited reach. It is often challenging or impossible to identify the creator of a harmful deep fake, and they could be located outside the United States” (Chesney, Citron).

The internet is vast, and it is highly difficult to regulate every single piece of information that is put out into the digital world. While there still may be no true technological solution to the issue of malicious deepfakes, staying educated and letting others know about the impact they have is one of the few effective mitigation methods to date. Algorithms used to detect deepfakes could still have a breakthrough as it is still early in the process, but it still is incredibly valuable to have the awareness and skepticism about our media intake moving further into the age of AI advancements.

In conclusion, the evolution of AI and deepfake technology has potential to do great things for society. There are already life-changing technologies being created to help assist those who have sensory disabilities or cannot speak with their own voice. The ever-evolving nature that AI holds means that the possibilities are endless. Like a developing human, AI and deepfake technology will evolve into an even more intelligent and incomprehensible part of our world. The meshing of reality and technology is going to become more and more distinct in our worlds media and this may create distrust and uncertainty within our society. The deepfake issue is more than a question of the ethics and morals behind them, it encourages us to question the way this technology may manipulate and change our society, for the better or for the worst. Our world should not underestimate the power that AI and deepfakes will have on our society. Only time will tell how humanity will manage the issue of deepfakes, but it starts with creating awareness and keeping up to date with this quickly evolving technology.

Works Cited

Allyn, Bobby. "Deepfake Video of Zelenskyy Could Be 'Tip of the Iceberg' in Info War, Experts Warn." NPR, NPR, 17 Mar. 2022, https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia.

Chesney, Robert, and Danielle K. Citron. "Disinformation on Steroids: The Threat of Deep Fakes." Council on Foreign Relations, Council on Foreign Relations, 16 Oct. 2018, https://www.cfr.org/report/deep-fake-disinformation-steroids.

Eisikovits, Nir. "The Slippery Slope of Using AI and Deepfakes to Bring History to Life." The Conversation, 25 Feb. 2022, https://theconversation.com/the-slippery-slope-of-using-ai-and-deepfakes-to-bring-history-to-life-166464.

Engler, Alex. "Fighting Deepfakes When Detection Fails." Brookings, Brookings, 14 Nov. 2019, https://www.brookings.edu/research/fighting-deepfakes-when-detection-fails/.

Groh, Matt. "Detect DeepFakes: How to Counteract Misinformation Created by AI." n.d. MIT Media Lab, https://www.media.mit.edu/projects/detect-fakes/overview/.

Jaiman, Ashish. "Positive Use Cases of Synthetic Media (Aka Deepfakes)." Medium, Towards Data Science, 6 Jan. 2022, https://towardsdatascience.com/positive-use-cases-of-deepfakes-49f510056387.

Miller, Joshua Rhett. "Deepfake Video of Zelensky Telling Ukrainians to Surrender Removed from Social Platforms." Nypost.com, New York Post, 17 Mar. 2022, https://nypost.com/2022/03/17/deepfake-video-shows-volodymyr-zelensky-telling-ukrainians-to-surrender/?utm_source=url_sitebuttons&amp;utm_medium=site%20buttons&amp;utm_campaign=site%20buttons. Accessed 9 Apr. 2022.

Somers, Meredith. “Deepfakes, Explained.” MIT Sloan, 21 July 2020, https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

Sosa, Jared, director. *You Won’t Believe What Obama Says In This Video! YouTube,* uploaded by BuzzFeedVideo, 17 Apr. 2018, https://www.youtube.com/watch?v=cQ54GDm1eL0

Vaccari, Cristian, and Andrew Chadwick. “Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News.” Social Media + Society, Jan. 2020, doi:10.1177/2056305120903408.