Analysis on The Usage of Deepfakes

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Brandon Ongtingco

Spring 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Joshua Earle, Department of Engineering and Society

INTRO

A deepfake is a video of a person in which their face or body has been digitally altered so that they appear as someone else. Originally these deepfakes began on the internet as a simple meme, such as putting someone's face on various cartoon characters and then having them dance as if it was the actual person dancing. Later on, it further evolved to putting people's faces on real people such as Donald Trump, or Joe Biden and impersonating them as a simple joke.

However, deepfakes have the potential to cause big social damages. Deepfakes have many potential legal implications such as forgery, and impersonation of others. These can lead to many different cases such as swaying people to a particular point as if it was portrayed by someone else. Overtime the constant evolution of technology and editing has made deepfakes a lot easier to create and could see further development to be able to use them to damage someone's public image or to assist in one's own self-gain. Eventually deepfakes could potentially be used for more serious crimes such as bypassing voice recognition to gain access to private files or causing potential identification errors. Within this report, I ask the question "Are deepfakes tools that could be used for good, or only for evil?" I examine this by looking at multiple resources and research papers regarding both positive and negative aspects of deepfakes in a general case, and then make further judgments based upon my findings.

CREATION OF DEEPFAKES

Deepfakes are formed using techniques stemming from both Artificial Intelligence and Machine Learning. According to the Library of Congress [8], a primary method of deepfake creation is the usage of "generative adversarial networks (GANs)." These GANs use two different machine learning systems known as the "generator" and the "discriminator." The

“generator” is used to help create counterfeit data such as photos, audio recordings, or videos that replicate many qualities of the original source material. Meanwhile, the “discriminator” is used to determine the difference between the original data and the newly created data. By using these two networks, deepfakes are able to be created. The generator is able to create more realistic deep fakes, and the discriminator tries to break down the content and tell them apart. This process would continue until the deepfake is the closest to perfection that it can get. The very first instance of a deepfake occurred in 2017 when a reddit user by the name of “Deepfakes” posted code allowing other users to create their own deepfakes. In 2018 another reddit user, this one anonymous, further adjusted the code to be implemented in an app that’s known as “FakeApp” which made the creation of deepfakes more accessible to the public [4]. Since then, deepfakes have become more mainstream on social media platforms with a majority of them being seen as parodies and various other jokes. Even nowadays deepfakes are being used in media, such as a deepfake of Luke Skywalker in The Mandalorian. However, there are a lot of different potential usage for deepfakes.

THE DEEPFAKE PROBLEM

The absolute nature of deepfakes is still up in the air, but there’s a significant problem with deepfakes. Deepfakes have the potential to cause distrust in the validity of evidence. Take for instance in a court of law, the defense submits a video of the defendant being in a location other than the crime scene, but the whole scene was actually just created using a deepfake. This could severely impact the verdict of the trial without anyone ever knowing that the evidence was forged. As Riana Pfefferkorn says “Modern jurors have been raised to believe that the camera does not lie [9].” In 2018, researchers demonstrated that at the time, a lot of police body cameras could have their footage be remotely downloaded, digitally manipulated, and then re-uploaded, thus making video evidence not fully credible. One example was in a California

appeals court case “People v. Beckley,” where a photo from MySpace had not been properly proven to real due to there being a lack of testimony on the contents of the image. Another example is a Colorado state appeals court case, “People v. Gonzales,” where the judge said “While software has made it easy for laypeople to manipulate recordings, ‘the fact that the falsification of electronic recordings is always possible does not, in our view, justify restrictive rules of authentication that much be applied in every case when there is no colorable claim of alteration’.[9]” While this could be seen as both a positive or a negative depending on whether or not the video was a deepfake or not, it still causes a lot of potential problems regarding trust and legitimacy of many court cases and other matters that were resolved with digital media.

NEGATIVE ASPECTS OF DEEPFAKES

Deepfakes are internet creations that may have started out as harmless, but have many potential damaging effects if used maliciously. According to Robert Chesney and Danielle Keats Citron [4], “there are eight potential harms to society resulting from the use of deep fakes.” They list these harms as: distortion of democratic discourse, manipulation of elections, eroding trust in institutions, exacerbating social divisions, undermining public safety, undermining diplomacy, jeopardizing national security, and undermining journalism. In regards to manipulation of elections, there are two primary threats from deepfakes: “the use of deepfakes to alter voter preferences, and the impact of deepfakes on trust generally in elections and democratic institutions [1].” In some instances, deepfakes can be used to strongly influence the votes of the people. One example being if a candidate that heavily advocated against drugs was suddenly seen taking drugs due to a video create from a deepfake. Some examples of this involve videos by the Republican Party attacking Nancy Pelosi by slowing down videos of her to make her seem drunk, or videos of Joe Biden during the 2020 presidential election making false statements [1]. As deepfakes have improved, it is easy for people to believe that deepfakes are

real people. In 2020, a study found that “approximately 15% of viewers in a controlled trial believed a deepfake of Obama was real” [1]. These deepfakes could cause quite a few potential political problems such as voters being slightly influenced to change their votes, or by potentially bringing down a candidate’s credibility and forcing them to withdraw.

Another big issue with deepfakes is the use of deepfakes in revenge porn. This is where women can have their faces edited into a pornographic film, making it seem as if it was the actual person performing the acts in the film. One particular issue that arises from this is that it further paints women as sexual objects. “By treating women’s faces as a digital resource to be edited onto sexual bodies by artificial intelligence, [deepfakes] reinforces the idea that women exist as sexual objects” [2]. With photoshop, and now deepfakes, women’s bodies have constantly been digitally adjusted which can not only cause these women as individuals to have a bad public image, but also can leave them with unique scars. With these sexual deepfakes, women can feel humiliated and violated through someone making a sexual video of them, regardless of it being real or not.

From these instances we can see that there is a severe problem of deepfakes being misused. From court cases, to influencing people’s politics, to getting revenge, there are many potential ways to use deepfakes in a negative manner. Deepfakes are something that can potentially ruin the lives and prestige of many different people.

POSITIVE ASPECTS OF DEEPFAKES

While deepfakes are mainly known for their negatives, they have positive impacts as well. One positive aspect of deepfakes is to help recreate media. While stated previously that deepfakes are used to create nefarious media such as revenge porn, people also use it to do things such as replace actors who have passed away, or help recreate people’s voices [5]. One

example is in “Star Wars A New Hope” with the late Carrier Fisher’s character Princess Leia being created with visual effects and deepfakes. This helped the overall quality of the movie since it allowed the studio to keep previous footage of Princess Leia while still being able to make more to finish up her character arc. Another example can be seen during “A 2019 global malaria awareness campaign featuring David Beckham” [5]. In this awareness campaign deepfakes were used to make David Beckham appear bilingual both audibly and visually by making the audio translate into different languages smoothly and by having the face change properly based on the facial and mouth movements needed.

Another positive usage of deepfakes is to help improve human interaction with one another. One example can be seen in virtual chat worlds online, such as VR Chat or Metaverse. In this case deepfakes have helped improve proper facial tracking and proper audio sound to make interactions online seem more real. This can also help people revisualize themselves. A transgender person, for example, could use deepfakes to visualize who they truly want to be in the virtual world. With the constant improvement of deepfakes we could see many changes and improvements in the digital landscape which will help human interactions feel more personal and real.

Deepfakes can also help in terms of marketing. While this can also be seen as a negative there is a lot of positives in regards to making things more appealing for consumers. With the increase in technology, deepfakes can help create “photorealistic images of scenes generated through this technology are especially useful in advertising and marketing content creation (Picazo & Moreno-Gil, 2019)” [6]. This can also be utilized in exhibits. One example is seen in the Dali Lives exhibition in the Dali museum in St. Petersburg. In this exhibit, the museum uses deepfakes to “recreate an immersive visitor interaction and learning about Salvad Dali from the renowned artist himself” [6]. With the constant improvement of deepfakes and

holograms we can see these types of instances occur more and more to help make user experiences more personal and enjoyable.

One final usage of deepfakes can be to help medical diagnoses. New technologies are created all the time to help make life easier as well as help develop society, this can also be seen in the medical field. One particular technology that has seen constant improvements is artificial intelligence, the primary technique used to create deepfakes. In this paper by Thambawita [7] they look at electrocardiograms (ECGs) with both the standard methods and with the usage of deepfakes. For context, an electrocardiogram "is a voltage time series that reflects the electric currents within the heart" [7]. These are primarily used to detect cardiac diseases. In this study they created two different GANs in order to generate a bunch of ECGs to determine which should be classified as "normal" ECGs. The main purpose of doing so was to see if it was possible to one day have technology create their own patterns of recognition for doctors to rely on which can help maintain the privacy of their patients by just having the machines analyze the data instead of sole individuals. The results of this study were that around 81% of the ECGs created were labeled as Normal ECGs using the new method of identification. This new method has an increased identification rate from the previous 75%. This shows that with enough improvement in technology, we can rely more on artificial intelligence technology to help perform certain operations and be used in training.

Despite deepfakes being known for mainly negative connotations, there are a significant number of positive possibilities, too. Throughout time, with the rise of deepfakes and continued technological advancement, there can be many good usages for deepfakes and artificial intelligence technology. With constant advancement and with the right purpose, we could see deepfakes being used more often for recreational use, or for training our service workers of the future.

METHODS TO SOLVE THE DEEPFAKE PROBLEM

With the rise of deepfakes it's important to be able to have countermeasures to make it possible to properly determine what is real and what is fake. It is also important to make sure that deepfakes are not used in a professional setting like political campaigns, popular media sources, like the news, or magazines. Currently, there are two bills that are used in targeting deepfakes [10]. The first one is known as "The Malicious Deep Fake Prohibition Act of 2018" which was introduced in the Judiciary Committee. This bill offered fines and/or up to two years of imprisonment for anyone who created a deepfake with the intent to distribute, or someone who distributes will full fledge knowledge of the content at hand being a deepfake. The bill also offered fines and/or 10 years in jail for politically oriented deepfake aimed at any form of Federal, State, local, or Tribal government agency. One final argument is that in response to deepfakes being protected under the first amendment, "no person shall be held liable under this section for any activity protected by the First Amendment to the Constitution of the United States" However, this bill was not passed.

The second bill, introduced on June 12th, 2019 is known as "Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019." This bill required all deepfakes to have disclosures based upon the type of content used. It also created a task force under the Department of Homeland Security which reports to Congress to determine whether or not an incidence of fraud has occurred due to a deepfake. Under this bill, a criminal could go to jail for up to five years and/or be fined if the deepfake is used to 'humiliate' or 'harass' an individual in a sexual way [10]. Along with elections, this bill also extended the penalties described to "a foreign power, or agent thereof," that violates the disclosure in an attempt to influence any sort of public election [10]. In regards to the First Amendment, this bill authorizes the United States Attorney General to waive the requirements of the law as they

deemed necessary where “the producer can demonstrate compliance with this section would impede their ability to engage in otherwise lawful activities protected by the First Amendment of the Constitution [10].” Unfortunately, this bill was not passed, however it was a step forward in the right direction.

Another solution would be to identify a way to recognize whether something is fake. Deepfakes are created by Artificial Intelligence, and we can use the same technology to reverse the process in order to determine whether something is real or fake. There are a few ways to be able to spot a deepfake. One of them is to observe the eyes of the person in a video. One flaw in deepfakes is that the rhythm of blinking eyes in deepfakes tends to be either too fast or too slow, causing a substantial identification trait [11]. Another weakness can be seen in the teeth or hair. The current technology behind deepfakes have trouble capturing accurate details. Certain key differences, such as facial structure can stand out among other instances. One final weakness deals with the background. Sometimes the creator will try to get the face to match, but will not focus on the things in the background. Examples could be a discolored background, or sound dropping off at certain points due to the technology not properly adjusting to the background as well as it adjusts to the person. In addition, we can inform people about deepfakes to ensure that they become more aware about the potential dangers of online videos.

FURTHER ANALYSIS OF DEEPFAKES

In terms of an ethical standpoint of deepfakes, there’s a slight amount of confusion. With regards to some ways that people use it, it can be seen as unethical and an invasion of privacy since they are using other people’s faces and voices for their own needs/entertainment. On another hand however, we could say that a majority of people already have a public presence online in some capacity, whether it’s through social media, the news, or just having accounts on various websites. In terms of developing actual technology and future use, I would say that this

could potentially help us more with automation. As stated in the positive aspects, the concepts of deep fakes could be used and developed further in other cases such as being used as training regimes. While there is a significant problem to deepfakes, I would say that it is something that is worth contributing to further development. Despite potential forgeries of media, it has been shown that deepfakes can also help a lot of people with their daily lives. There are many other ways that deepfakes could help people connect with each other and help them enjoy their overall lives instead of breaking them down. While the ways of identifying a deepfake are not fool-proof, it is a good way to start the argument that deepfakes are not something to fully fear. I do believe that based upon our current generation deepfakes are something that we are slowly becoming more aware of over time, and something that we will eventually become more inclined to recognize.

CONCLUSION

Deepfakes are a recently developing technology that has so much potential. As we continue further development, we begin to grasp more about the how deepfakes can be seen as both beneficial and malicious. In time deepfakes will become something that becomes indistinguishable from real people, and computer-generated media. We need to be able to properly moderate how we use new founded technologies and also be able to properly demonstrate all the potential hidden effects of technology before the general public comes to the conclusion that something is only used for jokes. If we are able to solve the deepfake issue, I believe that we can further develop techniques of artificial intelligence and machine learning to help improve our daily lives. This can already be seen partially in the usage of deepfakes to communicate with people online. While there is still a long way to go, it is possible that deepfakes could make long-distance communication feel more personal and real. At the same time, with the way things currently are and how easy it is to make one, it is also something that

could be dangerous and ruin credibility of people who are unknown on the topic. Deepfakes are truly something to be both admired and feared at the same time.

CITATIONS

[1] RAY, ANDREW. "DISINFORMATION, DEEPFAKES AND DEMOCRACIES: THE NEED FOR LEGISLATIVE REFORM." University of New South Wales Law Journal, vol. 44, no. 3, 1 Jul. 2021, pp. 983 - 1013.

[2] Gosse, Chandell, and Jacquelyn Burkell. "Politics and Porn: How News Media Characterizes Problems Presented by Deepfakes." Critical Studies In Media Communication, vol. 37, no. 5, 1 Dec. 2020, pp. 497 - 511.

[3] Hodge Jr., Samuel D.. "DON'T ALWAYS BELIEVE WHAT YOU SEE: SHALLOWFAKE AND DEEPFAKE MEDIA HAS ALTERED THE PERCEPTION OF REALITY." Hofstra Law Review, vol. 50, no. 1, 1 Oct. 2021, pp. 51 - 80.

[4] Ice, J. (2019, January 1). Defamatory Political Deepfakes and the First Amendment. Case Western Reserve Law Review, 70(2), 417 - 456.

[5] Mika Westerlund. "The Emergence of Deepfake Technology: A Review." Technology Innovation Management Review, vol. 9, no. 11, 1 Nov. 2019, pp. 40 - 53.

[6] Kwok, Andrei O. J., and Sharon G. M. Koh. "Deepfake: a Social Construction of Technology Perspective." Current Issues In Tourism, vol. 24, no. 13, 1 Jul. 2021, pp. 1798 - 1802.

[7] Thambawita, Vajira, et al. "DeepFake Electrocardiograms Using Generative Adversarial Networks Are the Beginning of the End for Privacy Issues In Medicine." Scientific Reports, vol. 11, no. 1, 9 Nov. 2021, pp. 1 - 8.

[8] Sayler, K. M., Harris, L. A., & Library of Congress (issuing body) (2020). Deep Fakes and National Security ([Library of Congress public ed.). Washington, D.C.: Congressional Research Service.

[9] PFEFFERKORN, RIANA. "'DEEPFAKES' IN THE COURTROOM." Boston University Public Interest Law Journal, vol. 29, no. 2, 1 Jul. 2020, pp. 245 - 276.

[10] Bodi, M. (2021, January 1). The First Amendment Implications of Regulating Political Deepfakes. Rutgers Computer and Technology Law Journal, 47(1), 143 - 172.

[11] Lamphere, C. (2021, September 1). Deepfakes Revisited: How Transformed Technology Poses New Challenges. Online Searcher, 45(5), 33 - 35.