Deepfakes, Our Future or Our Downfall?

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Brandon Ongtingco

November 1, 2021

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.

ADVISORS

STS Advisor: Joshua Earle, Department of Engineering and Society

Technical Writing Advisor: Rosanne Vrugtman PhD, CS Department
Technical Advisor: Daniel Graham, CS Department

INTRO

For my thesis, I will explore the concepts of A.I. in social media. Specifically, I want to look into what are known as “deepfakes”. 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. Deepfakes began on the internet as a meme, however deepfakes aren’t just ways to get a laugh. Deepfakes have many potential legal implications such as forgery and impersonation of others that could damage one’s career and social image. There are many deepfakes for famous cartoon characters such as characters from SpongeBob, but there are also deepfakes for real people too such as Donald Trump. Even something small can lead to big social damages in one’s relationship, it could cause people to believe things that were supposedly said from other people even though they are deepfakes. While harmless, deepfakes can still potentially be used to sway people to a particular point if it portrays the proper person. With the proper editing and technological advancement, we could see deepfakes be used constantly throughout our life span in order to damage someone’s public image, or to assist one in their own self-gain. Eventually deepfakes could be used for more than just audio, and could one day potentially be used to gain access into private files through voice recognition or identification errors. Deepfakes are a current rising threat of society and seem to be the next big potential crime dealing with A.I.

Deepfakes

Deep fakes are created from techniques using both Artificial Intelligence and Machine Learning. According to the Library of Congress [2], one main method of creating deep fakes is the usage of “generative adversarial networks (GANs)”. GANs use two different Machine Learning systems, the “generator”, and the discriminator. The “generator” is used to create counterfeit data such as photos, audio recordings, or video that replicate many qualities of the original source material. The “discriminator” is used to determine the difference between the

original data and the newly created data. With a combination of these two networks, the generator is able to create more and more realistic deep fakes, while the discriminator tries to break down the content and tell them apart. The first instance of a deepfake occurred in 2017 when a reddit account with the name “Deepfakes” posted code allowing other uses to create their own deepfakes. Later in 2018, another unnamed reddit user adjusted the code to be implemented in an app known as “FakeApp” which then made the ability to create deepfakes even more accessible to the general public [4]. Since then, the rise of deep fakes have taken over social media and other platforms by storm. A majority of them can be seen as parodies and various other jokes.

Negative Aspects of Deepfakes

Deepfakes are a creation of the internet that can seem harmless but have potentially very damaging effects if used maliciously. According to Chesney and Citron [4], “there are eight potential harms to society resulting from the use of deep fakes”. They list them 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.” We can see the beginning of potential deepfake problems in a court of law starting back in 2010 court case, People v. Beckley. In this court case Albert Jerome Beckley and Darrell Amont Finn were convicted on a count of first-degree murder. During the trial, the prosecutor attempted to submit a photo from MySpace.com regarding Finn and Beckley’s involvement in a gang by having one of them display gang signs [5]. In the trial it was stated that “a photograph is a ‘writing’ and ‘authentication of a writing is required before it may be received in evidence’”. Although the jury was convinced that the correct person was identified in said photograph, there was no definite evidence, and as such, the photograph was not allowed to be submitted as acceptable evidence. This is just the beginning of technology being used in a court of law. As we progress more and more, I feel that either defendants or prosecutors will continue to use deepfakes as potential forged evidence and, in trials where the

jury is easily persuaded, be able to condemn someone that is innocent. Although more and more laws are coming to light where each piece of evidence is thoroughly examined, it will overtime become more and more difficult to differentiate between what is real and what is not.

Another example of negative aspects of deepfakes can be seen in “revenge porn”. “Revenge porn” is a term added to the Merriam-Webster dictionary in 2016 of April and is defined as “sexually explicit images of a person posted online without that person’s consent especially as a form of revenge or harassment.” However, revenge porn is seen as something distributed that is actually genuine as opposed to content that was doctored [7]. Pornographic deepfakes however can be created by anyone, and can be used to create a man’s fantasy. Both of these types of pornography can leave victims with severe emotional and psychological damage. One big threat with deepfake porn is that it currently may not always be considered a crime. In a majority of situations with deepfakes, these videos are seen as harmless jokes rather than ones that pose serious threats. One example is when Steve Buscemi’s face was place on Jennifer Lawrence’s body in which she discussed her favorite “Real Housewife” [7]. This also leads to another potential problem of the true victim. In these deepfake pornography, there is an issue regarding who is the real victim since the person whose face is deepfaked could face a similar amount of emotional harm as someone whose body is used. This can lead to very complicated situations and have a significant number of victims compared to revenge porn.

Positive Aspects of Deepfakes

However, deepfakes aren’t always negative. Deepfakes also have a positive impact as working as training tools. Deepfakes can actually help with training and learning the proper techniques. Some examples can be in art, forensics, medical, and security. A few positive usages of deepfakes can be seen in very common situations. One example in a general case is how the usage of deepfakes can create “synthetic media” that can create “personalized assistive navigation apps for pedestrian travel” which can make things more familiar to some users. This can be seen in Microsoft’s “Seeing.ai” and Google’s “Lookout” [6]. There’s a few

other medical usages that can be very helpful for others. One example can be seen in patients with Amyotrophic Lateral Sclerosis which can cause nerves dealing with muscle functionality to breakdown. One potential positive of deepfakes is that with continuous development of deepfakes, it can help these patients become able to talk with their loved ones after losing the ability to use their voice [6].

Key Texts

For my research I will continue analyzing different articles as well as looking further into the usages of Artificial Intelligence and Machine Learning in our daily lives. By doing so I will be able to gauge a better understanding of the technology presented and will be able to further determine whether or not deepfakes can play both a positive and a negative role in society.

The first important text that I will continue to reference is "Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act." by R.A. Delfino. This text is important because it gives some thoughts and deepful insights of what is known as "revenge porn" and how deepfakes can be used in it. This article also talks about the legal problems with deepfakes in pornography and I believe that is something that will help lead me to more topics to discuss within my paper.

The second important text is "Positive use cases of deepfakes" by A. Jaimin. This article is not as descriptive as other citations; however, it opens a lot of different perspectives and avenues for positive usages for deepfakes. While there are a lot of supporting devices and small amounts of evidence used in this paper, it also allows me to gain more ideas on different potential usages for deepfakes so I can examine the possibility of some of these options.

A third important text is **"**Defamatory Political Deepfakes and the First Amendment. Case Western Reserve Law Review**"** by J. Ice. This article is useful because it goes into a more legal aspect of deepfakes and how they could be used court. With the court cases and potential scenarios presented in this article, I will be able to determine more potential issues where deepfakes can occur and also situations where more laws may need to be adjusted or

implemented in order to counteract the current problem with deepfakes. From this article, I hope to be able to gain a better understanding of the problem with deepfakes in a court setting and how it compares to other doctored evidence.

Conclusion

In short, Deepfakes are a recent 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 undistinguishable from real people, and computer-generated media. In time 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.

Citations

[1] Blankenship, R. J. (Ed.) (2021). Deep Fakes, Fake News, and Misinformation in Online Teaching and Learning Technologies. Hershey, Pennsylvania: Information Science Reference.

[2] 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.

[3] Fairfield, J. A. T. (2021). Runaway Technology: Can Law Keep Up?. Cambridge University Press.

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

[5] *People v. Beckley, 110 cal. rptr. 3D 362, 185 cal. app. 4th 509*. CourtListener. (n.d.). Retrieved November 4, 2021, from https://www.courtlistener.com/opinion/2265148/people-v-beckley/.

[6] Jaiman, A. (2021, October 5). *Positive use cases of deepfakes*. Medium. Retrieved November 4, 2021, from https://towardsdatascience.com/positive-use-cases-of-deepfakes-49f510056387.

[7] Delfino, R. A. (2019, January 1). Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act. Fordham Law Review, 88(3), 887 - 938.