Deepfakes, Our Future or Our Downfall?

A Technical Report submitted to the Department of Computer Science

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

Technical Project Team Members

Brandon Ongtingco

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

Technical Writing Advisor: Rosanne Vrugtman PhD Department of Computer Science

Technical Advisor: Daniel G. Graham PhD Department of Computer Science.

Deepfakes, Our Future or Our Downfall?

CS4991 Capstone Report, 2021

Brandon Ongtingco
Computer Science
University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia, USA
bmo4aa@virginia.edu

ABSTRACT

Deepfakes are a rising threat and may be the next big potential crime in the field of Artificial Intelligence. According to the Oxford Dictionary, a deepfake is “a video of a person in which their face or body has been digitally altered so that the person appears to be someone else.” Originally deepfakes appeared on the internet as a meme trend to get a laugh. However, deepfakes have potential legal implications, including forgery and impersonation. While they can be harmless, deepfakes can still potentially be used to sway people to a particular point of view if they are believed to be true. With the proper editing and technological advancement, we could see deepfakes used in ways that damage someone’s public image, or to promote fraudulence. For instance, deepfakes could be used to gain access into private files through voice recognition or can be used in modern day training modules to improve performance of new recruits in specific fields.

1 Introduction

Deepfakes are “video forgeries that make people appear to do or say things they didn’t. [1]” The term deepfakes was created in 2017 by a reddit user named “deepfakes” who began to share code and videos that showed celebrities’ faces swapped onto the bodies of actors and actresses in pornographic videos. There are many deepfakes for famous cartoon characters such as characters from Spongebob, but there are also some for real people too, like Donald Trump. Deepfakes started as something that was done solely for laughs, but the real question is, will deepfakes become a technological disaster?

2 Related Works

According to the Library of Congress, one main method of creating deep fakes is the usage of “generative adversarial networks” also known as GANs [1]. GANs use two different Machine Learning systems known as the “generator” and the “discriminator”. The “generator” creates the counterfeit data that replicates aspects of the original source material, while the “discriminator” is used to determine the difference between the original data and the new data. By combining these two networks, the “generator” helps to create more and more realistic fakes, while the “discriminator” constantly tries to break the content in order to differentiate between the real content and the fake content.

The nature of deepfakes is one that’s not solely set-in stone but significantly leans down towards one side. The general consensus paints deepfakes as something that is will continue to be the downfall of society and technology. One instance is in 2017 of August where the University of Washington researchers released a video of what appeared to be Barack Obama talking about terrorism, fatherhood, and job creation. However, this video was nothing more than a deepfake created with the usage of Machine Learning [2]. This action can be seen as a serious crime, one that is not easily distinguished from a real person talking, which makes the situation more complicated and even more difficult to troubleshoot.

However, some people say that deepfakes can also lead to a positive influence. The concept of deepfakes is just a small step in users becoming more familiar with the capabilities of both machine learning and artificial intelligence. With continued development and time, deepfakes can also be helpful in many different usages. A few examples could be helping improve cancer diagnoses, helping cartographers and astronomers to create more maps of the universe, and could help more with homeland security and identification [3].

Due to the uncertainty of the future, the nature of deepfakes is in question. While people want to remain hopeful and have the confidence that it will be a benefit to society, there's just not enough evidence and development, there is not enough information yet to determine whether the net effect is likely to be positive or negative.

3 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 video 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.

Another potential problem is impersonation. Impersonation can lead to many various problems and influence groups of people in a negative way. Deepfakes of political leaders and celebrities can be used to widely spread a message that may trick some people into following certain values or beliefs that the actual individual does not support. This can cause potential riots and confusion and affect the public's trust in online videos and news channels.

A positive case of deepfakes can cause the start of growth in using artificial intelligence and machine learning in more general areas of our lives. Deepfakes use generators and differentiators to be able to tell the difference between what is real and what is not. For example, development of deepfakes can be used to help with identifying people, not just deceiving them. If an individual goes to a federal building where security clearance is needed, then there needs to be a full proof system to properly identify that the individual in question has proper clearance. With deepfakes, a system can use a picture of the actual person as reference and then cross-reference it to the person presenting themselves in order to determine if it is someone that has access to the building or not. With these instances, it becomes more difficult to determine if deepfakes are more harmful or more helpful in future development.

4 Solutions to 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 [4]. The first one is known as "The Malicious Deep Fake Prohibition Act of 2018" which was introduced to the Committee on the Judiciary. 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 [4]. 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 [4]. 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 [4]."

Another solution would be to be able 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 [5]. 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.

5 Conclusions

Deepfakes are constantly evolving and are likely to become more widespread over time. While it is unclear what direction deepfakes will go, it is definitely something that will be important to be aware of in the future. Deepfakes could either be used negatively by other people and countries to try and confuse the general public and influence their views, or instead be used to help train and assist the future generation in identifying trouble areas. While deepfakes are constantly being developed and created, they are not foolproof and can still be differentiated. It is important to be able to recognize the traits of a deepfake, and also for the general public to become more familiar with them, and the laws against this misuse of technology. We need to become more and more aware of both the dangers, and benefits of deepfakes and decide if they can be used for good or for evil.

6 Future Work

Two CS classes that contribute the most to deepfakes are Artificial Intelligence and Machine Learning. Artificial Intelligence is obviously very relevant to deepfakes because it is the main source of face swapping and the technology that causes the faked parts of people move in a video. Machine Learning is equally as important because it allows for tracking and recognition of the initial subject to then be able to differentiate between other people and causes the system to be able to create a deepfake of them. I believe that our current classes provide a solid foundation on how these technologies work, but are not as thorough or specific enough to get into unique types of creations such as deepfakes.

One way to contribute to learning in both the Computer Science department and in the public eye would be to have some sort of activity or coding challenge to spot out the differences between deepfakes, and real photos/videos. It could be where there is a database of pictures of both real people, and deepfakes, where students could code or determine how/why a particular image is a deepfake or not. They could then later on try to create their own deepfake using machine learning and artificial intelligence in order to further help them with recognizing deepfakes and what they could look like. This would help students gain more hands-on experience and give them an opportunity to work with generative adversarial networks.

This similar activity could be done with people in public as well to show them how convincing deepfakes could look and how easy and common they can be. I think the best way to have people learn about deepfakes are activities like these that can help people know what to look out for. I feel that for the general public it may be slightly difficult to ask them to try and code something to counter-act the deepfakes, but for people in the computer science field, I feel that asking students and workers to try and code a deepfake could help with more recognition in them and thus could potentially help create more countermeasures against them.

REFERENCES

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

[2] Gerstner, E. (2020, January 1). Face/Off: "DeepFake" Face Swaps and Privacy Laws. Defense Counsel Journal, 87(1), 1 - 14.

[3] Clauser, M. A. (2020, January 1). Defending the Technology Behind Deepfakes. Ivey Business Journal, 2 - 4.

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

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