Deepfakes - a boon or a threat: A Review

Naman Mishra[1], Priyank Signhal[2], Shakti Kundu[3]
[1,2,3]*FOECS, Teerthanker Mahaveer University, Moradabad*
[1]namanmishra007@gmail.com
[2]priyanksinghal1@gmail.com
[3]shaktikundu@gmail.com

***Abstract*— Camera never lies used to be a popular saying but with the rise Image editing tools like Photoshop that quote was put into question a long back but Videos, videos were still solemn if you have a video of any person doing/saying something it has to be true. But all this was put into question with the rise of the Deepfake technology. With the availability of free to use apps, it is becoming very easy to create seamless Deepfakes. All you need is some relevant data and the software does all the heavy lifting. Deepfake technology if used with caution can change lives and save millions of dollars, but until now it seems to be doing more harm than good. It has been used in defamation, pornography, blackmail, and extortion. Therefore, we must be aware as to where we stand with the Deepfake technology and all it has the potential to do. Which is why through this paper, we tried to review this relatively new yet infinitely powerful tech and examine where we stand in a world where Deepfakes exist.**

***Keywords*— Deepfakes, fake videos, Threats to leaders, scams, healthcare and Virtual Humans.**

I. INTRODUCTION

Manipulation of images, audio and video is not a new concept, the film industry has been doing it for decades but it has always been a time-consuming activity that can cost studios millions of dollars while having mixed results. This is where AI-driven Deepfake technology can help, Yes Deepfakes, a technology which many prominent personalities in the literary world have said, they see as a threat Democracy. Deepfake technology seems to be a big concern as it has the potential to generate fake news and give power to scammers. It at the same time also has the potential to become the next big thing in the world of Artificial Intelligence as it is extremely easy to use. Basically, anyone with a computer, an internet connection, and general knowledge can create a Deepfake. Many researchers have been able to create Deepfake videos which seemed almost real. What we found even more interesting is what the technology behind Deepfakes has the potential to do.

A. What are Deepfakes?

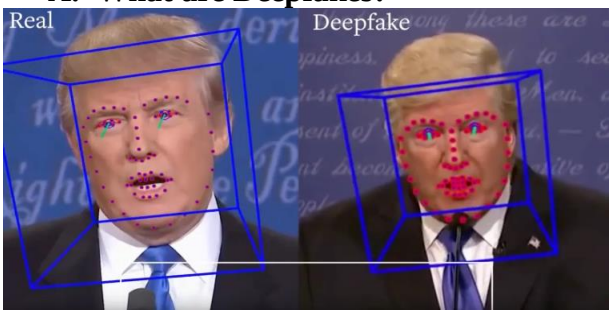

Fig. 1 Deepfake image of Donald Trump

In the term Deepfake, the word deep comes from the involvement of deep learning and the fake often refer to forged videos or audio. Deepfake are usually trained using the Generative Adversarial Networks(GANs), or autoencoders. Any video or audio which has been altered entirely or partially to give the impression that something happened, which in reality, did not happen is an example of Deepfake [1 , 2]. To create a Deepfake you need some image data of the person you want to add into a particular video as dataset A and then images of the face that will be replaced as Dataset B, these Data sets will be used to train our network to morph the two faces. The Faceswap.py software which is powered by Tensorflow, Keras, and Python can be used by normal users to create good Deepfakes [20]. In simple words, Deepfake can be any manipulated video that is created by an Artificial Intelligence and the intent of which is to appear real.

II. IMPACT OF DEEPFAKES

Even though we are anticipating that the technology behind Deepfakes will be responsible for some great advancements in the world of Video Editing and Artificial Audio generation, the sad truth is until now this powerful technology has been associated with scammers, hacks, and Trolls. The technology has affected some in a very adverse way be it scamming companies out of millions using a fake voice module or using pornographic videos to blackmail women. What is even more worrisome is that some experts think this is the surface of what this tech can do in the hands of the wrong people. But all is not so grey as companies like Google, Adobe and numerous Silicon Valley start-ups are trying to come up with ways, the Deepfake technology can be positively utilized by the masses.

A. Content Creation

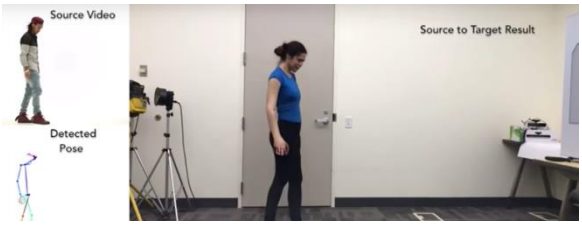

Fig. 2 Humen.ai at work

The truth is Deepfake technology is a gift to individual creators, who don't have huge budgets and teams for making and editing videos. The company Humen.ai is the best example of how Deepfake technology will one day become the epicenter of content generation, Humen is known for creating Deepfakes of people with no rhythm dancing perfectly, shown in fig 2 [3]. The amount of funny content already created with Deepfakes is enough of a sign that this technology will become popular with the creator community as more user-friendly Apps become available.

B. Cybercrimes and Deepfakes

Audio Deepfake has been used by scammers to get people to transfer money by pretending to be someone else over the phone. According to a report published on The Verge, "A UK energy company's chief executive was tricked into wiring €200,000 (or about 220,000 USD) to a Hungarian supplier because he believed his boss was instructing him to do so[4,5]. But the energy company's insurance firm, Euler Hermes Group SA, told the WSJ that a clever AI-equipped fraudster was using Deepfake software to mimic the voice of the executive and demand his underling pay him within the hour." This is just one of many instances and as the service gets more easily accessible it will only rise. Many startups across the globe are working towards the same tech and with the introduction of free to use Apps/Websites like LyreBird, FakeApp, Deepnude, and Faceswap.py, the bad seed of the world will continue to take advantage of these advancements.

C. Politics

Fig.3 Sample (A) Real (B) Comedien (C) Faceswap

Many believe that Deepfake will be used by politicians for their gain, as one wrong move and careers of many leaders can end not only this it is feared that corrupt politicians may use Deepfakes as an excuse to deny real evidence against them [6]. With the US Presidential election just around the corner experts are

worried that Deepfakes will be playing a prominent role to swing voters towards certain candidates [7].

III. HOW TO SPOT DEEPFAKES

[8]Deepfake detection is currently the priority of tech giants like Facebook, Google who are taking the approach of fighting AI created Deepfakes with their own respective AI's. This is becoming a major concern for various Government agencies and lawmakers all over the world as well who are turning to academic institutions to develop measures to detect Deepfakes [9]. We have a few methods currently used to spot Deepfakes

A. Human Moderators

The idea is simple companies hire human moderators who on the basis of few inconsistencies like inconsistent head poses, unnatural blinking, and inconsistent skin tones, decide whether a video is real or has been altered [10,11]. This is what was the basis of Deepfake detection for some time but as Deepfake AI Algorithms become better they are able to trick moderators so we needed more modern methods to spot Deepfakes.

B. Recurrent Neural Networks

This is as of yet is an experimental concept that has not been applied on a large scale. David Guera and Edward J. Delp of Purdue University used a temporal-aware system to automatically detect Deepfake videos and found a large collection of manipulated videos have shown that using a simple convolutional LSTM structure, they were able to accurately predict if a video had been subject to manipulation or not with as little as 2 seconds of video data [12]. The idea behind the method seems to be what can be the key to spot Deepfakes with less effort.

C. Browser plugins

We see a lot of content on our screen on a daily basis and it's hard to detect what's real and what's not this is where these plugins come, once added they will scan each page and tag any AI-generated content, Reality Defended by AI Foundation and Sursafe created by Ash Barth and Rohan Phadte are a great examples of how normal people can be alerted about Deepfakes [16, 17].

IV. FUTURE OF DEEPFAKES

Deepfakes came into mainstream attention in late 2017 when people's social media feed was filled with famous movies where faces of some actors were replaced with that of Nicolas cage. This made people think of Deepfakes as a technology that will save possibly millions of dollars and will bring advancements that could have never been thought of [18]. We have only being to understand how we can utilize this technology

A. The Film Industry

Fig. 4 Editing by professionals vs Deepfake made by fan

Warner Brothers reportedly spent over 25m Dollars and to remove some facial hair from Henry Cavil's face and even then the result came was horrible, which made a fan so furious that he trained deep learning algorithms with images of Cavil and ended up with a better result with-in 24 hours. Now imagine what a trained professional would be able to with this technology. We are yet to see the utilization of this tech by any major Film studios but sooner or later this technology will most probably replace all its more expensive counterparts [13, 19].

B. Educational Applications

This technology can be used to make learning more fun for future generations, think about it a video/hologram of Steve Jobs us about how we can create better products or Nicola Tesla

himself teaching about his accomplishments. This is possible if we put efforts of Deepfake towards this. This is how we can work towards making the prominent figures of our time immortal for generations to come.

C. Virtual Humans

Fig. 5 Human avatars created at Samsung labs

Every single person in fig.5 is a Digital avatar created by Samsung and not a real human being, So, Deepfake can be used to create specific Virtual humans for specific people and can act as an assistant, friend or as compadre. Taking the idea of virtual assistants such as Google Assistant or Cortana to the next level [14] .

D. Healthcare

Suggesting use of Machine Learning and AI in healthcare is often considered controversial as it will affect patient privacy [6,15]. But in research conducted by NVIDIA, MGH & BWH Center for Clinical Data Science and the Mayo Clinic it was found that using Deepfake Application such as GANs can potentially create medical images based on 10% real data which could be used to train Algorithms, yet the technology is very new and its too early for any use on a larger level.

Conclusion

Deepfake technology is now already among us which means that no matter what is done to control it, advancements in this field will be happening. What we need is for our world leaders to come together and create rules and regulations which will help in regulating the use of Deepfake technology, so that we can utilize the tech for all the good it has the potential to do. As for detection of Deepfakes, we think developing an AI while also maintaining human involvement in the process of detecting Deepfakes will be our best bet in this fight against Deepfakes. As for the future of the technology the possibilities are limitless, what made us most excited is the recent research finding of how tech Deepfake technology has the potential to contribute towards advancements in healthcare. Even with all the good and bad that comes with Deepfakes we are excited for what the future holds for Deepfake technology. We hope with this paper we were able to draw some of your attention to all that is happening with Deepfakes and were able to remove some stigma around the topic.

References

[1] The Emergence of Deepfake Technology: A Review November 2019 Mika Westerlund https://timreview.ca/article/1282

[2] Deep Learning for Deepfakes Creation and Detection by Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen,Duc Thanh Nguyen and Saeid Nahavandi

[3] https://www.humen.ai/ Humen visited on 14/02/2019

[4] Deepfakes: A Looming Challenge for Privacy, Democracy, and National Security by BOBBY CHESNEY AND DANIELLE CITRON

[5] https://www.theverge.com/2019/9/5/20851248/deepfakes-ai-fake-audio-phone-calls-thieves-trick-companies-stealing-money

[6] Protecting World Leaders Against Deepfakes Shruti Agarwal and Hany Farid University of California, Berkeley Berkeley CA, USA

[7] https://observer.com/2019/06/deepfakes-combat-2020-elections/

[8] .https://www.nytimes.com/2019/11/24/technology/tech-companies-deepfakes.html

[9] https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

[10] https://www.wired.com/story/ai-deepfakes-cant-save-us-duped/
[11] EXPOSING DeepfakeS USING INCONSISTENT HEAD POSES Xin Yang , Yuezun Li and Siwei Lyu
[12] Deepfake Video Detection Using Recurrent Neural Networks David Guera Edward J. Delp
[13] https://www.nydailynews.com/entertainment/movies/justice-league-reshoots-superman-mustache-removal-cost-25m-article-1.3351896
[14] https://thenextweb.com/artificial-intelligence/2020/01/07/samsung-is-working-on-lifelike-ai-powered-avatars-to-fill-in-for-humans/
[15] Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine Andriole, Mark Michalski
[16] https://www.wired.com/story/surfsafe-browser-extension-save-you-from-fake-photos/
[17] https://aifoundation.com/responsibility/ visited on 13/02/2020
[18] https://www.weforum.org/agenda/2019/11/advantages-of-artificial-intelligence/
[19] https://analyticsindiamag.com/can-deepfakes-be-used-for-the-good-of-humanity/
[20] https://faceswap.dev/ for Faceswap.py information visited on 10/02/2020