HW7, Part 1

Submitted By: Nafis Neehal

RIN: 661990881

Title: Deepfake – A Good Samaritan or A New God of Destruction?

1. Chosen Application and Technical Problems:

I chose to work with Deepfake - a very fascinating yet dangerous piece of technology! Deepfakes refer to manipulated audio-visual contents which are produced by sophisticated AI algorithms that yield fabricated images, sounds or videos that appear to be very realistic. For producing Deepfakes, GANs (Generative Adversarial Networks) are typically used, where the Generator module tries to produce fake images which tends to be as close to real images as possible, and the Discriminator module tries to differentiate this fake image with the real one and try to establish it as a fake[1].

The applications of Deepfake has, so far, created more negative social and moral impacts[2] rather than facing technical problems that developers might face while building it. These issues will be addressed in a nutshell in the following sections.

2. What else might this technology be used for:

Extreme level of audio/visual content manipulation can be done using this technology, which comprises (but not limited to) making hate-speech videos of popular political figure to defame them, making pornographic videos of celebrities and many more - if we are to emphasize only on the negative side. But at the end of the day, this is just a piece of technology (a very powerful one), and how we use it is entirely upon us. And, I choose to focus on the bright side today.

Some Positive applications of Deepfake can be / that currently exists[3-5] –

- Making videos of historically popular figures giving speech about different issues for solely **educational purpose**. Now, Albert Einstein can be seen on the 4k Screen of your TV taking an intense class on Quantum Mechanics. In a nutshell, Deepfake can become a game changing technology in the upcoming years for the billion-dollar story telling industry in Hollywood.
- **Synthesia** has a commercial product that uses Deepfake technology under the hood to do automated and convincing dubbing through automated facial re-animation
- **Adobe VoCo** is working on generating speech to text with artificial voice narration. That day is not to far when we can make our own animated movie staying in home and “Morgan Freeman” will give voice in the movie.
- **Humen** creates Deepfake for dancing, where a person’s move, who is dancing, are transferred and superimposed on another person’s body and making that other person dancing with the same moves as the first person.
- **Respeecher** and **Replica.AI** provides voice mimicry accessible for non-tech persons where Deepfake technology is used under the hood.

3. **How would someone collect training data for this application?**
   Most of the training data for building Deepfake contents come from web-scraping. With almost 93 million of selfies taken everyday and uploaded into the internet, building dataset for Deepfake is not a hard task today.

4. **What Social and Ethical Issues that might come with this data collection?**
   I believe, the major issue would be here to collect these data from internet via simple web-scraping techniques without taking any kind of permission/consent whatsoever from the person whose image is being used in training.

5. **What are some positive and negative consequences of building and deploying this application?**
   Deepfake is more of a technology, both sides of which coin should be equally emphasized and how we apply these in building different applications is up to us. As discussed, the most prominent feature of Deepfake is the capability to generate very much real-like audio-visual contents. We can use this to spread positive propaganda about the issues like – climate change, disease control and many other alarming issues if we want to stay on the positive side. But on the contrary, this same technology can be used to fulfil destructive means by spreading criticism, racism, hatred, vulgarity and so on.
   One positive thing that the existence of Deepfake has instilled us is not to believe everything that we see on the internet. Some more positive example of positive uses Deepfake (apart from the ones mentioned in section 2) so far[8],
   - **"Malaria Must Die"** campaign: David Beckham's video was generated in 9 different languages to spread awareness about Malaria
   - **831 JFK** speeches were used for training to generate JFK's Dallas Speech for educational purpose by "CereProc" (before which he got assassinated

   Some negative obvious consequences of applying deepfake technology into building something harmful are[6] –

   - Spread negative propaganda
   - Gain political advantage over rivals
   - Reputation damage
   - Spread fear/anarchy
   - Ignite chaos in social platforms
   - Spread fraud, humiliation and misinformation

6. **Based on current technology should we build this application? Should we restrict its use if we do build it? Why and how?**
   As I have stated, Deepfake is a technology and both "Good" and "Evil" types of applications can be built on top it. We really neither can put a break on advancement of technology, nor should we do it. But what we can do is to choose how to use this piece of powerful technology to spread positive news, propaganda and happiness.
   There are already some measures taken to detect deepfake news[7], [9] –
   - University of Washington researchers has built an application called "Which Face Is Real" with tutorials for effectively spotting deepfake images

- A large-scale video dataset released by a private research organization for fogery detection is now being vividly used as a benchmark dataset for classical image forensic tasks
- Deeptrace, a startup combating deepfake cyberthreats, has published a detailed report on the current state and developments of deepfake in 2018
- Hany Farid at Dartmouth is developing an open-source project which can effectively identify political deepfake videos
- Siwei Lyu, in collaboration with DARPA, is developing open-source software to effectively detect and thwart the spread of deepfake videos
- Giants like Google has started working on detecting Deepfakes. According to their release of "AI Principles", they committed to develop best practices of AI. To do that, they released "Synthetic Speech Dataset" and "FaceForensicsBenchmark" has been released with multiple collaboration to reduce the harm the Deepfake technology can cause[10].

As concluding remarks, I believe if developed and applied carefully, prioritizing moral dilemmas and ethical point of views, Deepfake can become a powerful technology, a positive Samaritan to revolutionize the future world.

References:

1. https://www.popularmechanics.com/technology/security/a28691128/deepfake-technology/
2. https://www.cnbc.com/2019/10/14/what-is-deepfake-and-how-it-might-be-dangerous.html
3. https://techcrunch.com/2019/07/04/an-optimistic-view-of-deepfakes/
4. https://hackernoon.com/the-light-side-of-deepfakes-how-the-technology-can-be-used-for-good-4hr32pp
5. https://artificialintelligence-news.com/2019/08/05/dont-believe-your-eyes-exploring-the-positives-and-negatives-of-deepfakes/
6. https://medium.com/twentybn/deepfake-the-good-the-bad-and-the-ugly-8b261ecf0f52
7. https://cybersecurity.att.com/blogs/security-essentials/deepfakes-are-a-problem-whats-the-solution
8. https://www.forbes.com/sites/bernardmarr/2019/07/22/the-best-and-scariest-examples-of-ai-enabled-deepfakes/#15dbcd4a2eaf
9. https://medium.com/@jonathan_hui/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9
10. https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html