Cyber Security and Threats: Deepfakes Impacts and Risks

Shazah Ishtiaq
School of Technology
Asia Pacific University of Technology & Innovation
Kuala Lumpur, Malaysia
TP061756@mail.apu.edu.my

Nor Azlina Abd Rahman
Forensic and Cyber Security Research Centre
Asia Pacific University of Technology & Innovation
Kuala Lumpur, Malaysia
nor_azlina@apu.edu.my

Khalida Shajaratuddur Harun
School of Technology
Asia Pacific University of Technology & Innovation
Kuala Lumpur, Malaysia
khalida@staffemail.apu.edu.my

Abstract - Deep Fake is a technology which was invented by Ian Goodfellow, the deep Fake works with generative adversarial networks (GANs). It is an algorithm which classifies a data into creating images or generating them. In this scenario two GANs they try to trick each other into thinking that the image is real the usage of the actual image is very less, and GAN can create a different version of that person. There have been many controversies around this technology which has been affecting the society, e.g. The US election, Mark Zuckerberg etc. There have been several research conducted on finding methods to deduce the negative effects of deep fake. This assignment will focus on finding the solution to the deepfake attacks and what potential risks that can be seen in such attacks. The impact of this attack will also be elaborated and what kind of techniques can be used to help in recovering an organization.

Keywords—Component, formatting, style, styling, insert

I. INTRODUCTION

Deep fake has been very popular these days specially on social media platforms. The use of AI technology help in transplanting a person's face into another person's body. This software uses a variety of images and videos of someone's face, and it creates a map of it, this map then can be fitted into any video or image of the creator's choice. There have been similar softwares created but the most closely related to deep fake is an application called Re-Face, it is a less advanced version of deep fake. They do create similar images, but they are more life-less products (Worrall, 2021). The first ever deepfake video was found back in 2017, when a renowned actress's face was used on a porn actor. This can be taken as an example that how this technology can create videos of world leaders by using false speeches for their own agendas. This technology if further abused can create differences between countries or misunderstandings between religion or politics. There are many ways this technology can be misused which can create problems which can be impossible to solve (Balas, 2020). Ransomware is a type of malware in which the attacker gets hols of the information of the target by encrypting its files such as locking it or completely wiping the memory, until the ransom is paid for it. When these two things are combined together a disaster can take place. For example, the actor can create the deepfake video of the target and can ask for money or else (Rockit, 2021).

There are some advantages of deepfakes as well for example, for people who have lost their voice in a web series they can create voice of the actors without re-shooting the actual scene. But because the negative aspects weigh more than the positive aspect of deep fakes it has been an enemy to the society.

Generative adversarial network which is also known as GAN creates an artificial illustration of data that easily passes for a real data, which means that the main resolution of it is to create a struggle between the generator and the discriminator. GAN's architecture includes a generator which uses multi-dimensional input vector to create an image which is fake, and the discriminator then recognizes the image and determines if the image is real or fake (Zotov et al., 2020). It can be observed in the Fig.1 given below.

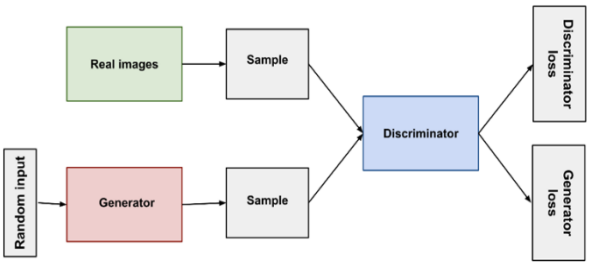

Fig. 1. GAN architecture (LIBBY, 2020)

The core components of deep fakes are the convolutional neural networks (CNN) which also performs a mutual function. The discriminator's task is to classify the end image which is directed from the generator or a dataset. But in this case, the use of CNN can be concluded by minimizing the resolution, integration of a video, passed through different layers, this results in an output which is then classified as an object in a particular class. In here, the generator has a different task, which includes the creation of a fake data. The convolutional networks are used in the structure of the generator, it uses multi-dimensional vector for creating an object. This implies that the object will then go through this layer and it will filter the data by lowering the sample's ratio, it will go through the layer once again and then create a new data (Zotov et al., 2020).

II. DEEPFAKE CREATION

Videos nowadays get tampered easily because of deepfakes commonality. Variety of online applications are available which can help individuals who have expert or beginner skills. These applications are created by using deep learning elements. The very first application was created by a Reddit user, he used the autoencoder and decoders blend in a single structure. The example of how these deepfakes is created can be seen below in the Fig 2.

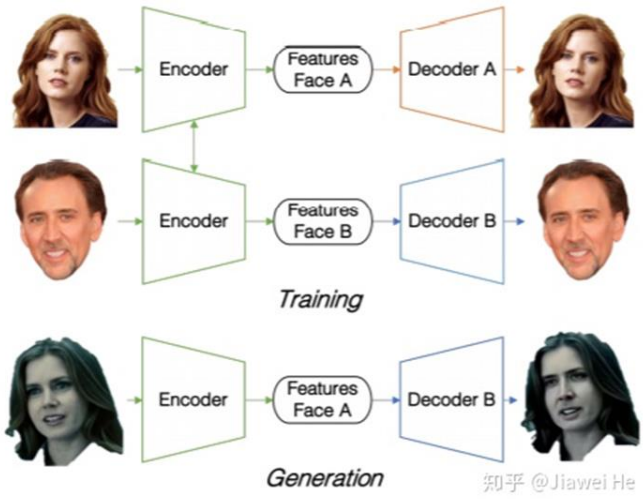

Fig. 2. How deep fake works (Pan et al., 2020)

In the method mentioned above states that there is an autoencoder present which extracts the features of the image and then uses decoder to reconstruct the image's face. Two encoder and decoder's sets are required between the source and target's images. Using this technique, the common encoder finds and learns the comparison between the two images, which is quite easy because faces have common structures such as nose, eyes, lips etc. In Fig 2. it can be clearly observed that the face A is connected to the decoder B to deconstruct the original face's features A to B(Pan et al., 2020).

III. TOOLS USED FOR DEEPFAKES

There are many tools available on the internet to create deepfakes. These tools doesn't have a good technology to give a perfect output but it does give an okay output. The details of the tools used for deepfakes are mentioned below:.

DFaker

It is tool which uses the Deepfakes techniques by reconstruction of the face, to achieve this reconstruction a tool is used, it is called DSSIM loss function Keras library-based implementation. It also uses GAN in its system as well (Pan et al., 2020).

Face swap

It is an online tool which performs deepfake process in each step, the importing of the videos and the final product of the video (RankRed, 2021). This tool uses the auto-encoder architecture, it consists of adversarial loss and perpetual loss (VGG-Face). This carries the CNN descriptors for implementation based on deep learning VGG(Oxford, n.d.). The technology used in this tool is GAN and MTCNN (Face detection tool uses python).Equations

Deep Face Lab

It is also popular amongst the deepfake softwares. This tool uses the neural networks to change the faces of a video or an image. This technology is accommodated on GitHub and it has generated several tutorials on the internet. The developers claim that 95% of the videos are created on this software (RankRed, 2021). The technology used in this software are GAN -er and MTS3FD

IV. DEEPFAKE DETECTION

Luckily, there have been various research which are going on for detecting the deep fakes. The benefits are due to the use of algorithms, they have the potential to detect cues within a video or an image which a human might have a hard time to find. There have been many detection methods which was found for both video and image detection. As they were several, they have been categorized in 2 major classes. The fake image detection and the fake video detection it can be seen in Fig 3.

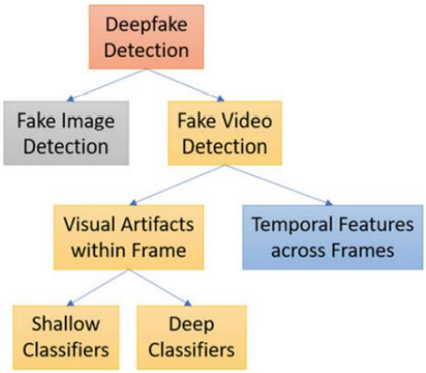

Fig. 3. Classification of Deepfake (Pan et al., 2020)

The video detection approaches are further elaborated: the visual artifacts for video frame-based and the temporal features in the frame-based methods. For the latter, machine learning methods such as deep learning is used, and for visual artifacts in a video can be performed by shallow or deep classifiers (Pan et al., 2020).

To find deep fakes a binary arrangement takes place in which the classifiers are used in between the reliable and interfered videos. To use this method a huge database is required of both the fake and real videos, so that the models can be trained accordingly. The discover methods for fake videos has been very limited whereas the number these videos are available. There was a research which was conducted by (Korshunov & Marcel, 2019)it produced a different type of data set of 598 videos based on the GAN technology with the help of the open source tool called Face swap-GAN. The videos which were collected were available online at VidTIMIT database, this database contained both high-quality and low-quality videos of deepfake. The videos were chosen based on how the facial features, actions of the mouth and blinking of the eyes can be seen. To test different deepfake videos this data base was used, the results of the test mentioned that face recognition supported VGG and Facenet which could not help in detecting the deepfakes successfully. According to the research the lip syncing and the quality of the image's metric with the use of SVM concluded high errors in the detection of the deepfakes. From this data it can be concluded that methods are required to detect deepfakes from the real ones (Korshunov & Marcel, 2019).

V. BUSINESS CONTINUITY PLAN

The case which will be looked on will be of a company who have been blackmailed by a threat actor on releasing the deepfake video of the CEO of the company and he blackmails to make this video viral. Nowadays attacks are so common in businesses that the need of assessing a businesses position after a certain attack is important.

For the business to prevent all type of disasters and it can recover from it a framework called Business Continuity Planning (BCP) can be used. This framework ensures that all the assets and the personnel are protected, so that they can work quickly after a disaster has struck. Fig. 4 elaborates what the BCP investigates. It involves all the risks and after effect of the operations in a company, this makes an important part of the company's risk management strategy. The risks can include anything such as a natural disaster or cyber-attacks. When all the risks are found then the planning is done:

- How the risks can affect the operations.
- Executing safeguards and practices to alleviate risks.
- Testing all the procedures to check if they are functioning.
- Reviewing all the processes which have taken place up to date.

Fig. 4. Business Continuity Plan

A) Business Impact Analysis

This stage will gather information which is required to develop strategies to overcome a disaster. The loss of business can be identified by using a risk assessment. Next the operations, they can also be damaged such as failure of a supplies or delay in the delivery of goods in a logistic company. Many scenarios are possible within a company which is why it is really important to consider BIA in a company (Ready, 2021). The basic theories about the BIA are a company relies on the functionality of another component. But in this framework some components require more attention in a middle of a disaster. There are no official standards that exists within BIA, the methods which are used in different organizations can be different (Carol, 2020). The following processes can be used in BIA stage they are as follows:

- Information Gathering
- Collected information's evaluation.
- Documentation of the findings
- Results to Seniors management

The conduction of detailed questionnaires to find the business processes which are critical, it can also involve the findings of critical resources and relationships as well. These information are very important for assessment of the potential impact of an event.

The collection of information can be collected by different ways such as interviews which are conducted in-person or surveys. To get more detailed information the follow-up interviews can also be conducted(Carol, 2020).

B) Misinformation and Impact

In the deepfakes case we can conclude that the information about the company can be called 'fake news'. Nowadays there are many journalists or article that can go viral if it is fake. As all the people are not journalists nor do they have the deduction skills to determine whether the news is fake or real, this can manipulate the direction of a market .(Rapoza, 2017) Back in 2013, there was a tweet which was posted by the Associated Press stated that there was an explosion, and the president of America Obama was injured. In that 2 minutes the stock markets had a loss of $125 billion value (Atkinson, 2019). This statement concludes that the impact if a fake news can be critical for a business. The following impacts can occur in the company:

- The stocks of the company can drop after the outbreak of the news.
- Customer's trust can be lost.
- It can be more prone to getting attacked by threat actors.
- The increase in the libel.
- It can affect the reputation of the brand.
- In the ransomware perspective the loss of confidential data.

C) Risk Management

Risk Management focuses on the prediction that can occur when there is something wrong in the plan. This management can help in reducing the impact of the attack to a tolerable level. Risk can be both positive or negative, it depends on the situation of an impact or an event .(APM, n.d.)According to the given case, the best practice to manage risks can be concluded by the process which as follows:

- Identifying the risk.
- Determine who might be harmed and how he/she can be harmed.
- Evaluate the risk and take measures.
- Document your findings.
- Review your assessment and update if it needs to (Lucid Content Team, 2020).

Fig. 5. Risk Assessment Plan

D) Defence Techniques

To combat against the fake news of a company it is important to effectively build the reputation of the company via using the press. The social media play a very large role in spreading of the fake news, so the best way to counter this issue can be concluded by the following methods:

a) Clear Headlines: Digital media has become dependent on the headlines, so if a deepfake video comes into picture. The first thing that needs to done is to play with the headlines without losing the authenticity of the company. The company needs to focus on a message which can portray the value of the article. It needs to be this good that the readers are impressed by it. The message should not be more than 10 words, short and simple with intriguing message.

b) Legitimate Third-Party Sources Utilization: Most of the information that comes from the social media are often from unreliable sources, which is why the company needs to work with the writer that can make sure that sources are not tangled with the company. For example, if a reader reads some article and it is directly linked to the company, they will automatically think the link is genuine. Approach articles which can notify people about the issue and make sure the sources are not biased.

c) Focus on the value of the brand: When focusing on the brand value, the most important thing is to be knowledgeable on how to use the brand power to inform all the people on a specific issue. In this case, the company should approach the writer and make them to get the attention of the media to the company by giving the most convincing pitch ever. Work hard together with writer to brainstorm every possible angle to educate the people on a specific topic.

d) Stay on top: No matter how the media has mentioned the name of the company, whether it is good or bad. The company needs o reach out to these writers and try to make good relations with them. The negative story about a company can go viral, make sure your part of the story is told clearly and fix the things accordingly (Olenski, 2018).

E) Awareness and detection

The increase of deepfakes has led companies like twitter and google to develop systems that can filter out deepfakes but for further protection from these types of attacks the things that an individual should point out are:

1) *The expressions:* A human being has around 42 individual muscles that makes them express a lot of expressions, if in a video the micro-expressions such as the movement of the eyebrows cannot be seen, it means that it can be an AI technology.

2) *Videos look unnatural cause of slow down:* There are websites which allows an individual to change the speed of the video. If a person is worried that he/she might be watching a deepfake video, the usage of that function can be helpful. It is available on YouTube as well.

3) *Blurry lines:* Check the pixels of the video or an image, most importantly around that person's jawline, forehead, and their hair strands (Nandi, 2021).

To deal with the deepfakes in day-to-day life, it is important to set a rule of principles which should be followed; They can include:

- Maintain a healthy suspicion. The company should make sure that all the content which is read amongst the employees should be checked thoroughly.
- The verification of the source is very important, nowadays there are many articles that comes from unreliable places which can create a lot of misunderstanding. It should be known that where the video came from and what's the purpose of it publishes.
- Education of such events is very important. All the employees should be aware of the implications of deepfakes. They should know how to distinguish the true information from the fake ones.
- Company should invest in deepfake awareness campaigns. So, it is known that deepfakes can create state level issues (Townsend, 2020).

VI. CONCLUSION

This assignment covered how dangerous it is to manipulate people by using the deepfakes technology. It was also elaborated that how these deepfakes are created and how someone can detect it with their naked eyes. As this technology is new, there can be a lot of problems that can occur even though the problems nowadays are just minor. It was also found out that it is a relatively new concept which is why the detection methods of it are very less. The impact of deepfakes can cause wars against people or countries which is why it is important to see this as a major threat to society and a company, which can occur in the near future. The case which was referred in this assignment was to give a risk assessments and impact in a company if it goes through similar attacks. Hopefully, these findings can be helpful to companies, and they implement this in their systems.

REFERENCES

APM. (n.d.). *What is risk management?* Retrieved February 12, 2022, from https://www.apm.org.uk/resources/what-is-project-management/what-is-risk-management/

Atkinson, C. (2019). *Fake news can cause 'irreversible damage' to companies — and sink their stock price*. https://www.nbcnews.com/business/business-news/fake-news-can-cause-irreversible-damage-companies-sink-their-stock-n995436

Balas, V. E. (2020). *Lecture Notes in Networks and Systems 146 Intelligent Computing and Networking.*

Carol, S. (2020). *What is a business impact analysis (BIA)?* Definition from WhatIs.Com, Tech Target. https://searchstorage.techtarget.com/definition/business-impact-analysis

Korshunov, P., & Marcel, S. (2019). Vulnerability assessment and detection of Deepfake videos. *2019 International Conference on Biometrics (ICB)*, 1–6. https://doi.org/10.1109/ICB45273.2019.8987375

LIBBY, K. (2020). *What Is a Deepfake? | Deepfake AI Technology Risks, Examples.* https://www.popularmechanics.com/technology/security/a28691128/deepfake-technology/

Lucid Content Team. (2020). *A Complete Guide to the Risk Assessment Process*. Available at: https://www.lucidchart.com/blog/risk-assessment-process

Nandi, S. K. (2021). *Deepfakes: On the Ethics of Contemporary Content Creation — Tekh Decoded.* https://tekhdecoded.com/deepfakes-when-contemporary-content-creation-goes-too-far/

Olenski, S. (2018). *5 Ways Brands Can Combat Fake News*. https://www.forbes.com/sites/steveolenski/2018/07/12/5-ways-brands-can-combat-fake-news/?sh=9393cac48f62

Oxford. (n.d.). *Visual Geometry Group - University of Oxford, VBB Face Descriptor*. Retrieved February 12, 2022, from https://www.robots.ox.ac.uk/~vgg/software/vgg_face/

Pan, D., Sun, L., Wang, R., Zhang, X., & Sinnott, R. O. (2020). Deepfake Detection through Deep Learning. *2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, 134–143. https://doi.org/10.1109/BDCAT50828.2020.00001

RankRed. (2021). *6 Best Deepfake Apps and Tools In 2021 - RankRed*. https://www.rankred.com/best-deepfake-apps-tools/

Rapoza, K. (2017). *Can 'Fake News' Impact The Stock Market?* . https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/?sh=4b754e062fac

Ready. (2021). *Business Impact Analysis*.

Rockit. (2021). *Deepfake ransomware explained - Rock IT: the Secure IT company*. https://rockit.cloud/2020/07/23/deepfake-ransomware-explained/

Townsend, C. (2020). *Deepfake Technology: Implications for the Future*. United States Cybersecurity Magazine. https://www.uscybersecurity.net/deepfake/

Worrall, W. (2021). *What Are Deepfakes And Why Are They Dangerous?* . https://hacked.com/what-are-deepfakes-and-why-are-they-dangerous/

Zotov, S., Dremliuga, R., Borshevnikov, A., & Krivosheeva, K. (2020). DeepFake Detection Algorithms: A Meta-Analysis. *2020 2nd Symposium on Signal Processing Systems*.