Fake News and Deepfakes: A Dangerous Threat for 21st Century Information Security

J. Botha[1], H. Pieterse[2]
Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa

[1]jbotha1@csir.co.za
[2] hpieterse@csir.co.za

Abstract: Fake news, often referred to as junk news or pseudo-news, is a form of yellow journalism or propaganda created with the purpose of distributing deliberate disinformation or false news using traditional print or online social media. Fake news has become a significant problem globally in the past few years. It has become common to find popular individuals and even members of the state using misinformation to influence individuals' actions whether consciously or sub-consciously. The latest trend is using Artificial Intelligence (AI) to create fake videos known as "deepfakes". Deepfake, a portmanteau of "deep learning" and "fake", is an artificial intelligence-based human image synthesis technique. It is used to combine and superimpose existing images and videos onto source images or videos using a machine learning technique called a "generative adversarial network" (GAN). The combination of the existing and source videos results in a fake video that shows a person or persons performing an action at an event that never occurred in reality. This paper provides an overview of the currently available creation and detection techniques to identify fake news and deepfakes. The outcome of this paper provides the reader with an adequate literature review that summarises the current state of fake news and deepfakes, with special attention given to the tools and technologies that can be used to both create and detect fake news or deepfake material.

Keywords: Artificial Intelligence, deepfake, detection, fake news, machine-learning.

1 Introduction

The earliest examples of fake news appeared during the 13th century BC when Rameses the Great spread lies and propaganda portraying the Battle of Kadesh as a stunning victory while the battle actually ended in a stalemate (Weir, 2009). Various other examples of fake news appeared throughout the centuries but the use of fake news became widespread during the 1900s, used as a form of propaganda during both the first and second world wars. However, the rise of the Internet during the late 90s led to the growth and accessibility of information (Posetti & Matthews, 2018), which in turn allowed fake news to grow exponentially. As of today, the Internet provides access to plenty of unwanted, untruthful and misleading information that can be produced by anyone, offering an ideal platform for propagation.

At present, fake news is a significant problem globally. It has become common to find popular individuals and even members of the state using misinformation to influence individuals' actions whether consciously or sub-consciously. Usually, fake news mimics news media content in form but not in the organisational process or intent. Fake news outlets, in turn, lack the news media's editorial norms and processes for ensuring the accuracy and credibility of information (Lazer, Baum & Benkler, 2018). Regardless, fake news still remains impactful as witnessed during the 2016 US election (Armstrong, 2016). Figure 1 shows the top five fake news stories shared on Facebook that gained the highest popularity during the 2016 US election.

In the era of manipulated images and the continuous presence of fake news articles, a new kind of technology is emerging, referred to as Deepfake. Deepfake is based on machine learning, usually deep learning techniques, which is used to produce or alter video to make it look like something happened that did not. The term "deepfake" is an amalgamation of both "deep learning" and "fake" and is a manipulation method that has been used in the studios of Hollywood for years (Davies, 2019). In the past, creating fake videos was an expensive task that required an extensive amount of skill, time, and money. However, today all it takes to create fake videos is a gaming laptop, an Internet connection, and a rudimentary knowledge of neural networks. There even exist applications that offer face-swapping in videos with a single click (Thaware & Agnihotri, 2018).

This paper provides a detailed literature review regarding fake news and deepfakes., with special attention given to the creation and detection of fake news and deepfakes. The outcome of this paper provides the reader with an adequate literature review that summarises the current state of fake news and deepfakes, allowing future research to explore options to design and develop new and improved detection techniques.

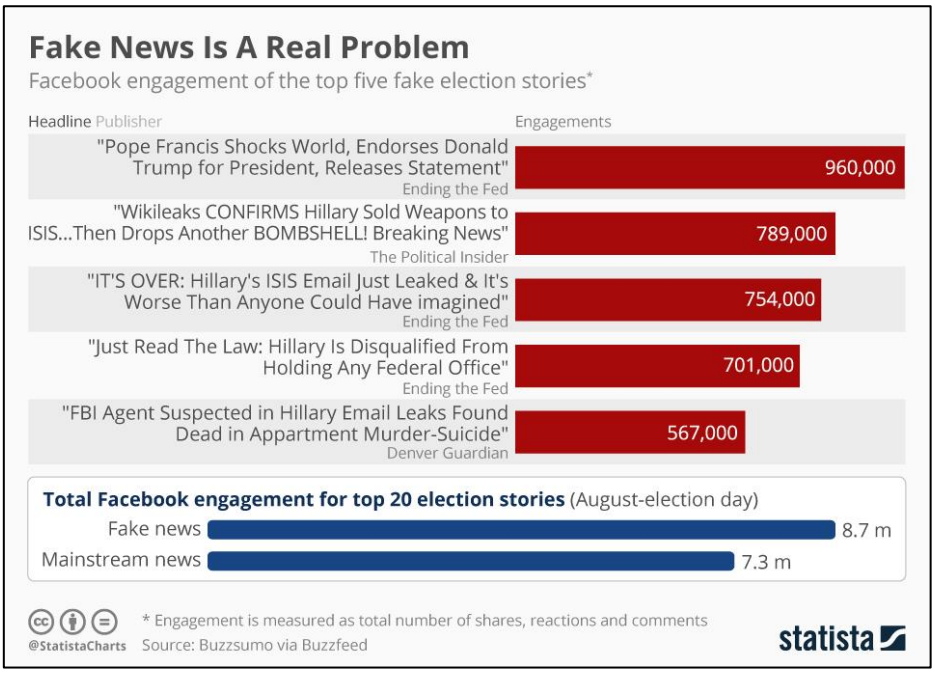

Figure 1. Top five fake news stories during the 2016 US election (Armstrong, 2016)

The remainder of this paper is structured as follows. Section 2 presents an overview of fake news and deepfakes while Section 3 focuses more specifically on the creation and uses of both fake news and deepfakes. In Section 4 the available technologies and tools to detect fake news and deepfakes are explored. Section 5 discusses the evolution of deepfakes as witnessed the past year and Section 6 concludes the paper.

2 Background

Fake news and the constant creation of deepfake videos remains a dangerous threat for 21st-century information security. There is, therefore, a need to continuously explore the potential and risks associated with fake news and deepfakes. This section presents a brief overview of fake news and the progression of deepfakes.

2.1 Fake News

The term “fake news” is not new and often refers to viral posts based on fictitious accounts made to look like news reports (Tandoc Jr, Lim & Ling, 2018). There are various types of fake news prevalent today. The most prominent types of fake news are the following (Lazer, et al, 2018):

- **Clickbait**: involves news stories that are purposely fabricated to gain more website visitors and increase advertising revenue for websites.
- **Satire/Parody**: news stories created purely for entertainment purposes with no intention to cause any harm, but may possibly fool the reader.
- **False connection/misleading headings**: include news stories or articles that contain genuine and accurate content but make use of misleading or sensationalist headlines.
- **Propaganda**: involves news stories that are deliberately created with the intended purpose to mislead audiences or promote a biased point of view or particular political cause or agenda.
- **Biased/Slanted news**: includes news stories that rely on the biases and beliefs of the readers.
- **Sloppy journalism/Error**: involves news stories constructed using unreliable information or untrusted sources that can mislead readers.
- **Manipulated news**: includes news stories that manipulate the content of genuine and factual stories to deceive the readers.
- **Fabricated news**: involves news stories containing 100% false content created to deceive and perform harm.
- **Sponsored content**: are news stories or advertising that is disguised as editorial content but can mislead the readers.

Two main motivations underlie the creation and development of fake news: financial and ideological. Firstly, outrageous and counterfeit stories often go viral simply because they are outrageous and provide content producers with clicks that

are convertible to advertising revenue. Secondly, other fake news providers produce false news stories to promote particular ideas or people that they favour, often by discrediting others (Allcott & Gentzkow, 2017). However, a new form of fake news has appeared.

2.2 Deepfakes: Fake News 2.0

The first appearance of Deepfake technology was on the Reddit social media platform by an anonymous user. Reddit is a platform comprised of more than a million smaller subreddits dedicated to different topics, which form small communities for content sharing and discussion. A user, only known as *u/deepfakes,* created such a community in November 2017. The community was aptly named *r/deepfakes*, and this is where the first face-swapping videos using the Deepfake algorithm started circulating (Fikse, 2018). As of today, these videos are easy to create and have become exceedingly sensible, convincing, and hard-to-recognise portrayals of genuine individuals doing or saying things they never actually stated or performed. As technology advances, the capacity to create false yet believable videos targeting celebrities, politicians, and governments, in general, will certainly increase. The implications of Deepfake technology can be detrimental to society or individual reputations and identities (Thaware & Agnihotri, 2018).

There have been multiple examples of Deepfake videos, most of which was created solely for entertainment purposes. However, a few key examples of the available Deepfake videos illustrate the potential power of the videos as a tool to conduct political or psychological operations.

One deepfake example is a video involving US President Donald Trump, encouraging Belgium to withdraw from the Paris climate agreement (BuzzFeed, 2018a). During the video, President Trump makes the statement: "*As you know, I had the balls to withdraw from the Paris climate, and so should you*". The video was published by a Belgian political party on both Twitter and Facebook but was eventually debunked by Lead Stories (Lytvynenko, 2018).

A second example shows a video starring Jordan Peele as former US President Barack Obama demonstrating the power and potential of fake news and face-swapping. The video was created using FakeApp and involved taking an original video of Barack Obama and pasting Jordan Peele's mouth into it. The final video took 56 hours to create (BuzzFeed, 2018b).

3 Fake News and Deepfake Creation and Uses

The improvements in computer technology and the current prevalence of Artificial Intelligence (AI) algorithms provide increasing opportunities for the creation of fake news and deepfakes. This section highlights technologies available to create fake news but more specifically focus on the methodology, tools, uses and limitations of deepfake videos.

3.1 Fake News

OpenAI, a non-profit AI research company, constructed a new natural language model called GPT-2. The GPT-2 model was trained to predict the next word in a sample of 40 gigabytes of Internet text. The result is a model that can "generate realistic and coherent continuations about a topic of their choosing" (Whittaker, 2019). The GPT-2 model was tasked with writing a response to the prompt, "*Recycling is good for the world, no, you could not be more wrong*", and the machine returned the following:

"*Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world's most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources.*" (Whittaker, 2019).

The ability of the GPT-2 model to generate long paragraphs of realistic text with greater coherence has caused some concern. The GPT-2 model can easily be used to generate fake news, impersonate individuals or create abusive spam comments on social media. Due to the concerns surrounding the capabilities of the GPT-2 model, OpenAI decided to only release a smaller version of the model to the public (Whittaker, 2019).

3.2 Deepfake

The availability of the Deepfake algorithm has allowed various automated Deepfake software tools to be created. Creation of a Deepfake video simply requires a selection of the images of the person's face being replaced and the person's face that is being superimposed. These following subsections describe the methodology to follow to create

deepfake videos, present a collection of tools to create deepfake videos and highlight the uses and limitations of the created deepfake videos.

3.2.1 Methodology

The availability and ease to generate AI-based fake videos have attracted a lot of attention the past year. At a high level, deepfake technology takes as input a video of a specific individual ("target") and outputs another video with the target's face replaced with that of another individual ("source"). The backbone of deepfake technology is deep learning neural networks trained on facial images to map the facial expressions of the source to the target. Deep learning algorithms interpret data features and their associated relationships using neural networks that evaluate the relevant information using several stages of data processing. For machine learning algorithms, the output will typically be a numerical value while deep learning algorithms can produce an output representing a score, element, test or sound (Badkar, 2017).

Deep learning neural networks, however, require a lot of source information, such as photos of the person who is the source or target of impersonation. The more images used to train a deepfake algorithm, the more realistic the digital impersonation will be (Lyu, 2018). Therefore, multiple software tools have emerged to streamline and simplify the process of creating deepfake videos.

3.4.2 Tools

The release of the deepfake algorithm on the Reddit platform led to the development of various software applications and web-based tools that can assist with the creation of deepfake videos. The current tools available to create deepfake videos are the following:

- DeepFaceLab (https://github.com/iperov/DeepFaceLab)
- FakeApp (https://www.malavida.com/en/soft/fakeapp/#gref)
- Deepfake Online (https://deepfakesapp.online)
- Deepfakes web (https://deepfakesweb.com)
- DeepFake.Me (https://www.deepfake.me)
- Faceswap (https://github.com/ModelIncubator/Deepfakes-faceswap)

3.2.3 Uses

The primary use of deepfake technology is to create false yet believable videos of individuals either portraying or saying something they never actually stated. Therefore, deepfake videos can become quite popular to promote negative political intent. Politicians are incredibly easy subjects for deepfake videos because the individuals are often standing face-forward on camera and there exists quite an amount of video material of the individual. From a negative perspective, deepfake videos can easily be used to create fake sex scandals for political figures, business leaders, celebrities and well-known people in the public eye (Davies, 2019). Available deepfake videos can be categorised into one of the following groups (Fikse, 2018):

- **Technology Demonstration**: include deepfake videos that are created as examples to demonstrate how the technology works.
- **Satirical/Meme**: involve deepfake videos that are humorous or mocking, created as a form of political or social commentary.
- **Pornographic**: involve deepfake videos that often carry the face of well-known celebrities on the body of pornographic actors.
- **Deceptive Deepfake:** fake videos made of political actors or an authority figure with the intention of creating a scandal.

3.2.4 Limitations

Technological advances simplify and allow for the creation of deepfake videos. However, even with all of the software and hardware components available today to create deepfake videos, there are still limitations to what can be accomplished. The key limitations associated with the creation of deepfake material are the following (Oberoi, 2018; Xu, Yang, Ye & Shen, 2018):

- An inadequate number of pictures will cause unsatisfactory learning results that in turn will produce Deepfake material of a low quality that can easily be detected.

- Adequate training set that is representative of the end goal, which can become a time-consuming process and require high computing resources.
- Constructing the model can be both expensive and time-consuming.
- Although existing models are available, these models are not free to use and require payment.
- The produced deepfake material can have the following limitation that may lead to detection:
  - Obvious differences in skin colour and other objective differences.
  - Obvious differences in illumination.
  - Faces with obvious edges, shadows, splicing or blurring effects.

4 Identification and Detection of Fake News and Deepfakes

The recent prevalence of fake news articles, and especially the growing concern of Deepfake videos, emphasise the need for technologies that can detect such false information.

4.1 Fake News Detection

The prevalence and potential impact of fake news, especially within the social, economic and political environments, emphasise the need to detect and identify fake news stories. The International Federation of Library Associations and Institutions (IFLA) published a summary in diagram form (see Figure 2) to assist people in recognising fake news (Blogs.ifla.org, 2017).

Various "Fact-Checking" websites exist that readers can access to evaluate news stories and eliminate fake news articles. The most renowned sites are the following:

- Snopes (www.snopes.com)
- PolitiFact (www.politifact.com)
- Fact-Check (www.factcheck.org)
- Channel-4-Fact-Check (www.channel4.com/news/factcheck)
- Africa-Check (https://africacheck.org)

Figure 2: Tips to identify fake news (Blogs.ifla.org, 2017)

The checks and "Fact-Checking" websites listed above offers initial steps to identify and potentially eliminate fake news. However, additional software tools are required to effectively detect fake news stories:

- The first tool is called fake-news-detection-with-deep-learning and relies on a state-of-the-art deep learning model to detect fake news based on the content of the news article (GitHub, 2019a). The training data consist of Snopes URLs that were randomly selected. The deep learning model was constructed using Tensorflow.
- The second tool is called Fake-News-Detection (GitHub, 2019c) and relies on various natural language processing techniques and machine learning algorithms to classify fake news articles. The data source used for training the program is the LIAR dataset (GitHub, 2019b).
- The third tool is FakeBuster (GitHub, 2019d), which also relies on natural language processing techniques and machine learning algorithms to detect fake news. However, instead of only relying on the content of the news article, FakeBuster also evaluates the source of the article. Once a source is labelled as a producer of fake news, it becomes possible to predict with high confidence that any future articles from that source will also be fake news.

4.2 Deepfake Detection

Although not as prevalent and widespread as fake news articles, the increasing popularity of Deepfake videos is a cause for concern. The sophistication behind the creation of Deepfake videos increases the difficulty of identifying and detecting such false videos. The remainder of this section explores manual and software-based detections solutions that can assist with the identification of deepfake videos.

4.2.1 Manual Detection

One of the most promising techniques to identify and detect Deepfake videos is to review the blinking of the individual within the deepfake video. Healthy adult humans blink somewhere between every two and ten seconds, and a single blink takes between one-tenth and four-tenths of a second to complete. Therefore, one will expect to see similar blinking in a video of a person talking but this is not the case in many deepfake videos (Lyu, 2018).

The reason for a lack of blinking in a deepfake video can be directly attributed to the algorithm used to create such videos. The training of the deepfake algorithm relies on facial images and very few such images show faces with eyes closed (Lyu, 2018). Therefore, the training data only includes faces with eyes open, which cause bias to the training data. The final trained model will have trouble understanding the action of a blinking eye and cannot produce a closed eye very well (Forsdick, 2019). The lack of blinking eyes in deepfake videos thus provides a simple but intuitive technique to detect deepfake material. It is possible to develop a method to detect when and if a person in a video blinks their eyes. The method will scan each frame of a video while detecting and locating the eyes automatically. Using another deep learning neural network makes it possible to determine if the detected eyes are open or closed by relying on the eye appearance, geometric features and eye movement (Lyu, 2018).

Another technique to detect deepfake material focusses on the inconsistencies between the angle of the head and the face. There are techniques available that can estimate the position the head is pointing to in a three-dimensional (3D) space on a two-dimensional (2D) video. With the creation of a deepfake video, the face is spliced into the video causing the head to point into a different direction to the camera. Therefore, the creators of the Deepfake video must perform a 2D transformation to wrap the face so that it matches the orientation of the head. The 2D transformation, however, introduces a lot of imperfections when the subject of the video looks away from the camera or changes the angle of the face position (Forsdick, 2019). Unfortunately, this particular detection technique fails to effectively detect Deepfake videos when the individual in the video always looks directly into the camera without changing the facial angle.

The third technique to identify and detect deepfake material relies on the imperfections caused during the process of creating fake videos. Imperfections caused by creating deepfake videos include the following (Hui, 2018):

- Double chins or ghost edges around the face.
- Over-blur comparing to other non-facial areas.
- A change of skin tone near the edge of the face.
- Double eyebrows or double edges on the face.
- The face is partially blocked by hands or other things.
- Flickering or blurring in the video.

These imperfections are produced because the creators of the deepfake videos cut corners to shorten the time required to create the videos, which can reduce the quality. The result causes the number of pixels on the face of the subject in

the original video to vary based on the distance from the camera and the size of the original image. The pictures containing the faces used to replace the original are generally a fixed size of 64x64 pixels or 128x128 pixels. To accommodate the variation, the fixed-size faces need to be transformed by enlarging, shrinking or rotating the picture to match the original video. These kinds of variations, when put together, will leave some imperfections such as an overly smooth face or a loss of detail. These imperfections can be detected by training a deep learning neural network to differentiate between the changes in detail in the facial region (Forsdick, 2019).

The final technique to identify and detect deepfake material is using blur detection. Blur detection is possible since the original face will have more levels of colour than the deepfake image when zoomed in. The main process of performing blur detection involves the following steps (Xu, et al, 2018):

1. Read in the video stream.
2. Locate the facial region using the Haar cascade classifier (machine learning object detection algorithm).
3. Convolve using a Laplacian operator.
4. Calculate variance for the face region.
5. Split into two cases: (1) known fake picture and (2) unknown video.

For the first case (1), compare the variances of the two facial regions. The picture with the bigger variance is the original while the picture with the smaller variance is the Deepfake image. The second case (2) will use the unknown video to locate another reference face region. For this new region, the variance will also be calculated. The next step will then calculate the ratio between these two selected face regions and compare the ratio to a threshold, which is established using a large collection of original pictures. Should the ratio be above the threshold, the video is original otherwise it is a Deepfake video (Xu, et al, 2018).

4.2.2 Software-based Detection

The first Deepfake detection solution, called Shallow (Shallow-ai.com, 2019), is a web-based application that makes use of a Keras convolutional neural network specialising in image recognition. The focus of the solution is to distinguish between real and fake videos with the intended purpose of protecting the reputation and integrity of anyone who can be affected by deepfake videos. Instead of relying on pre-trained weights used for classification of images, Shallow uses randomised weights to improve the accuracy of classifying deepfake videos.

Users can access the web interface and upload a video for processing. During the initial processing phase, Shallow will detect and extract the face crops available in the uploaded video. The user can then select 20 face crops for testing and run the images through the model. The model will perform the analysis of the video and report on the authenticity of the video (Shallow-ai.com, 2019).

The dataset used to construct and test the model for Shallow consist of four different categories: real training data, fake training data, real validation data and fake validation data. In order to ensure variety in both the training and validation sets, 30% of the images contain non-adult individuals while the remainder contains images of adults. The model was trained on two separate datasets and validated using an additional two datasets. The training was performed using 15613 different images with a ratio of 50/50 between deepfake images and real images. Validation was performed using 4872 different images with the same 50/50 ratio between deepfake images and real images. Based on the usage of the available training and validation datasets, the model was able to achieve 99% accuracy. However, it must be noted that the model was trained and validated using the available datasets and may not be representative of other datasets.

The second deepfake detection solution, called MesoNet (GitHub, 2019e), was developed to assist with the detection of face tampering in videos. The purpose of MesoNet is to provide a method to automatically and efficiently detect face tampering in videos and focusses specifically on two recent techniques used to generate hyper-realistic forged videos: Deepfake and Face2Face (transferal of image facial expression from a source to a target person) (Afchar, Nozick, Yamagishi & Echizen; 2018).

For both of the constructed models, the pre-trained weights are made available. The dataset used to construct and test the models for MesoNet consist of a training and validation set. The training was performed using 5111 forged images and 7250 real images. Validation was performed using 2998 forged images and 4259 real images. Both models demonstrated a very successful detection rate with more than 98% for Deepfake and 95% for Face2Face detection (Afchar, et. al.; 2018).

5 Evolution of Deepfake Videos

The technology used to create deepfake videos continues to evolve. Research conducted by scientists from Stanford University, Max Planck Institute for Informatics, Princeton University and Adobe Research demonstrated the ability to edit the text transcript of a video to add, delete, or change the words spoken by someone (Vincent, 2019).

In order to create such deepfake videos, scientists combine various techniques. The first technique involves scanning the target video to isolate phonemes (these are constituent sounds used to compose words such as "oo" and "fuh") spoken by the "target". Next, the scientists match these phonemes with corresponding visemes, which are facial expressions that accompany each sound. The final technique involves the creation of a 3D model of the lower half of the "subject's" face using the target video. All of these techniques, namely the phonemes, visemes and 3D model, are combined to construct the new footage that matches the text input. This result is then transferred onto the source video to create the final output (Vincent, 2019). Figure 3 presents the various techniques used to create a deepfake video that includes edited text transcript.

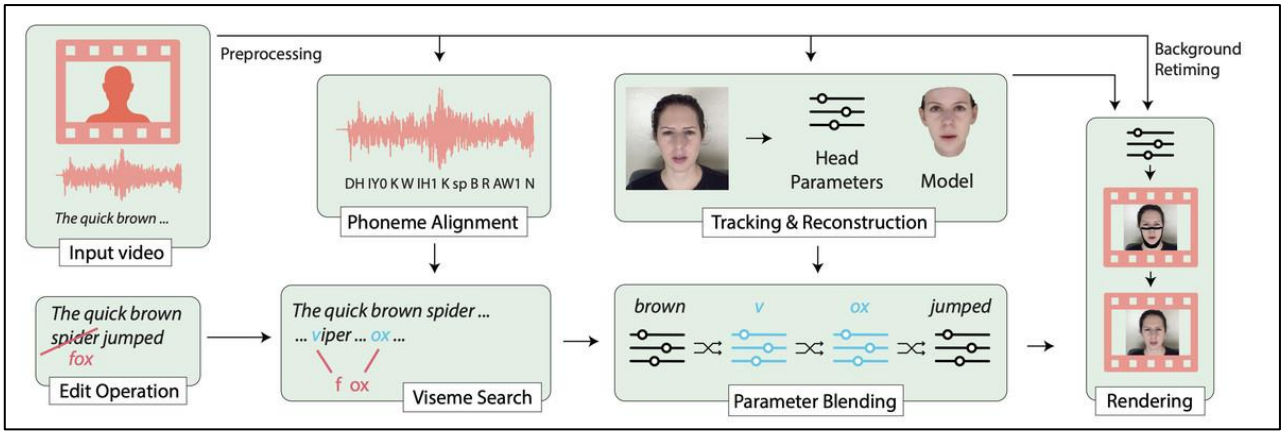

Figure 3: Illustration of the techniques used to edit a subject's speech (Vincent, 2019)

Although several successful demonstrations were presented, this research still has numerous limitations. The algorithms utilised only works using talking-head style videos and require approximately 40 minutes of input data. Furthermore, the edited text transcript cannot differ too much from the original source material (Vincent, 2019). This research is still in early development stages and while there are still various limitations, the potential misuse of this research is worrisome. Currently, this research is not available as consumer software but Adobe has already shared details on a similar prototype software name VoCo, which allows users to edit recordings of speech (Vincent, 2019).

Another example of the evolution of deepfake technology is the creation of a fake video using a single headshot. Researchers at Samsung's AI lab in Russia published a research paper that presented a technique that uses a portrait image to create a semi-realistic video (Kulp, 2019). Instead of relying on a neural network model with a large dataset of videos or photos, Samsung's AI Center created a method to train a model using a single photo and various landmark facial features (such as the shape of the face, eyes, mouth shape, etc.). Using the created model, it can map these features onto a photo to bring it to life (The Verge, 2019). The provided research and results are not without flaws, such as periodic flashes of distortion in facial movement, but offers a foundation to create deepfake videos without requiring a large dataset.

6 Conclusion

The focus of this research study were to investigate the current trends regarding fake material, such as fake news articles and Deepfake videos, and to explore techniques to create and detect such material. Fake news and the constant creation of deepfake videos remains a dangerous threat for 21st-century information security. There is a need to continuously explore the potential and risks associated with fake news and deepfakes. Therefore, this paper offered an adequate literature review that summarises the current state of fake news and deepfakes, with special attention given to the creation and detection techniques available to identify fake news and deepfakes. The purpose of this paper was to identify and review the currently available tools and techniques to create and identify fake news and deepfakes. The paper confirmed the availability of various software tools that currently exist to allow for the creation of fake news and deepfakes. Due to the potential societal implications of fake news and deepfakes, a requirement arose for the development of detection techniques that can pinpoint fake news and deepfakes. Although this paper identified and discussed a collection of detection techniques available to identify fake news and deepfakes, these techniques are still limited in the provided capability. Therefore, this paper provides the basis to support the development of new and

improved detection techniques for fake news and deepfakes. Future work will extend this research and explore the design and implementation of new detection techniques that can identify fake news and deepfakes.

References

Afchar, D., Nozick, V., Yamagishi, J. & Echizen, I. (2018). “MesoNet: a Compact Facial Video Forgery Detection Network,” in IEEE International Workshop on Information Forensics and Security, Hong Kong.

Allcott , H. & Gentzkow, M. (2017). “Social Media and Fake News in The 2016 Election.” Journal of Economic Perspectives 31 (2): 211–236.

Armstrong, M. (2016). Infographic: Fake News Is A Real Problem. [online] Statista Infographics. Available at: https://www.statista.com/chart/6795/fake-news-is-a-real-problem. [Accessed 24 Jul. 2019].

Badkar, A. and Badkar, A. (2017). Difference Between Machine Learning And Deep Learning. [online] iamwire. Available at: http://www.iamwire.com/2017/11/difference-between-machine-learning-and-deep-learning/169100 [Accessed 20 Oct. 2019].

Blogs.ifla.org. (2017). [online] Available at: http://blogs.ifla.org/lpa/files/2017/01/How-to-Spot-Fake-News-1.jpg [Accessed 28 Oct. 2019].

BuzzFeed. (2018a). A Belgian Political Party Is Circulating A Trump Deepfake Video. [online] Available at: https://www.buzzfeednews.com/article/janelytvynenko/a-belgian-political-party-just-published-a-deepfake-video [Accessed 3 June. 2019].

BuzzFeed. (2018b). How To Spot A DeepFake Like The Barack Obama-Jordan Peele Video. [online] Available at: https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed [Accessed 3 June. 2019].

Davies, S. (2019). “Deepfakes are the evolution of fake news and equally as dangerous,” 2019. Available: https://www.stedavies.com/deepfakes/. [Accessed 30 April 2019].

Fikse , T. D. (2018). “Imaging Deceptive Deepfakes - An ethnographic exploration of fake videos,” University of Oslo.

Forsdick , S. (2019). “Meet the team working to prevent the spread of next-gen fake news through 'deepfake' videos,” Compelo, 27 February 2019. [Online]. Available: https://www.compelo.com/what-are-deepfakes-detect/. [Accessed 17 May 2019].

GitHub. (2019a). nguyenvo09/fake_news_detection_deep_learning. [online] Available at: https://github.com/nguyenvo09/fake_news_detection_deep_learning [Accessed 29 Oct. 2019].

GitHub. (2019b). thiagorainmaker77/liar_dataset. [online] Available at: https://github.com/thiagorainmaker77/liar_dataset [Accessed 10 Oct. 2019].

GitHub. (2019c). nishitpatel01/Fake_News_Detection. [online] Available at: https://github.com/nishitpatel01/Fake_News_Detection [Accessed 29 Oct. 2019].

GitHub. (2019d). FakeNewsDetection/FakeBuster. [online] Available at: https://github.com/FakeNewsDetection/FakeBuster [Accessed 29 Oct. 2019].

GitHub. (2019e). DariusAf/MesoNet. [online] Available at: https://github.com/DariusAf/MesoNet [Accessed 29 Oct. 2019].

Hui, J. (2018). “How deep learning fakes videos (Deepfake) and how to detect it?,” Medium Corporation, 28 April 2018. [online] Available: https://medium.com/@jonathan_hui/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9. [Accessed 15 May 2019].

Kulp, P. (2019). “Samsung's AI Lab Can Create Fake Video Footage From a Single Headshot,” Adweek, 23 May 2019. [online] Available: https://www.adweek.com/digital/samsungs-ai-lab-can-create-fake-video-footage-from-a-single-headshot/. [Accessed 18 June 2019].

Lazer, D. M. J. , Baum, M. A. & Benkler, Y. (2018)“The science of fake news,” *Science,* vol. 359, no. 6380, pp. 1094-1096.

Lytvynenko, J. (2018). “A Belgian Political Party is Circulating A Trump Deepfake Video,” BuzzFeed News, 20 May 2018. [online] Available: https://www.buzzfeednews.com/article/janelytvynenko/a-belgian-political-party-just-published-a-deepfake-video. [Accessed 3 June 2019].

Lyu, S. (2018). “Detecting 'deepfake' videos in the blink of an eye,” The Conversation, 29 August 2018. [online] Available: https://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072. [Accessed 13 May 2019].

Oberoi, G. (2018). “Exploring DeepFakes,” KDnuggets, March 2018. [online] Available: https://www.kdnuggets.com/2018/03/exploring-deepfakes.html/2. [Accessed 13 May 2019].

Posetti, J. & Matthews, A. (2018). A short guide to the history of’fake news’ and disinformation. International Center for Journalists, pp. 1-19

Shallow-ai.com. (2019). Shallow. [online] Available at: http://shallow-ai.com/#/ [Accessed 29 Oct. 2019].

Tandoc Jr., E. C., Lim, Z. W. & Ling, R. (2018) Defining “Fake News”, Digital Journalism, 6:2, 137-153.

Thaware, V. Agnihotri, N. (2018). “AI Gone Rogue: Exterminating Deep Fakes Before They Cause Menace,” Blackhat Europe, London.

The Verge. (2019). Deepfakes are getting easier than ever to make, new research paper shows. [online] Available at: https://www.theverge.com/2019/5/23/18637373/deepfakes-samsung-ai-research-results-single-photo-algorithm [Accessed 28 Aug. 2019].

Vincent , J. (2019). “AI deepfakes are now as simple as typing whatever you want your subject to say,” The Verge, 10 June 2019. [online] Available: https://www.theverge.com/2019/6/10/18659432/deepfake-ai-fakes-tech-edit-video-by-typing-new-words. [Accessed 18 June 2019].

Weir , W. (2009). “History’s Greatest Lies” Beverly, Massachusetts: Fair Windows Pres.

Whittaker, Z. (2019). “OpenAI built a text generator so good, it's considered too dangerous to release,” Techcrunch, 17 February 2019. [online] Available: https://techcrunch.com/2019/02/17/openai-text-generator-dangerous/. [Accessed 17 Aug. 2019].

Xu, Z., Yang, Y., Ye, Y. & Shen, Y. (2018). “DeepFake Detection,” Github.