Advances In Multidisciplinary & Scientific Research
AIMS Research Journal
A Multidisciplinary & Interdisciplinary Journal

Volume 6, 2020 Edition

Situational Awareness of Deepfake Technology

Olebara Comfort Chinaza
Department of Computer Science
Faculty of Science
Imo State University,
Owerri, Imo State, Nigeria.
E-mail: chiiprime@gmail.com

ABSTRACTS

This work is a situational awareness reporting of the deepfake technology, which is a technology where multimedia components such as video. Image or audio of an individual is distorted with the aim of disinforming or misinforming the content viewers. The technology first appeared in November, 2017, when a social media user posted an algorithm that used AI to create fake videos. This paper will discuss areas where deepfake has been applied, some researches of how to create deepfake, and also some researches on how to detect deepfake. The study is mainly to report this new technology thereby making the society aware of its existence and watching out for its use.

Keywords: Deepfakes, Awareness, Security, Social Media, Online, Images, Videos, Users, Technology

Aims Research Journal Reference Format:
Olebara, C.C. (2020): Situational Awareness of Deepfake Technology". Advances in Multidisciplinary Research Journal. Vol. 6, Pp 153-157. Article DOI: dx.doi.org/10.22624/AIMS/V6N1P14. Available online at www.isteams.net/aimsjournal.

1. INTRODUCTION

Deepfake is the use of Artificial intelligence subset, machine learning, in the creation of fake video, audio, or image, usually for the purposes of entertainmentor misinformation. The term was coined from two words "deep" implying the useof deep learning models and "fake" informing that the resulting content is fake.

In deepfake, deep learning neural network simulation is applied to a large datasetand uses the dataset to learn what a particular face looks like when different fromvarious angles. The trained dataset can then be used to transpose the face into a target as Mask over the target. According to [1] two algorithms are involved in the deepfake creation: a generator algorithm used to model a source image, videoor audio (from a large dataset), then create a replicate version. The second algorithm known as discriminator works on the replica content and tries to detectanomaly in it. This short paper traverses cases of deepfake deployment in politics,with a view to creating situational awareness of this disruptive technology whichhas been declared a national security issue by the Defense Advanced Research Project (DARPA) [2].

Reportage of an interview with Dartmouth college computer Science Professor, Hany Farid, showed that DARPA started working on unravelling the use of AI in misinformation before the trending of deepfake, atechnology is capable of turning nations against one another especially with its presidents being credited with blaspheming speeches[2]. Nigeria, a third world country can be said to be digitally not in tune with currentglobal trends in technology. It is easy for few tech savvy Nigerians to take advantage of this and launch digitised political attacks that leave the politicians themselves bewildered. The power of visual and audio media cannot be overemphasized, especially in a society where majority of her citizens are belowglobally acceptable literacy level, making most people rely wholly on these two media.

While deepfake may imply the use of deep learning in manipulating media contents, synthesizing through other forms of media editing such as in audio spoofing would very much yield the same result and is currently covered by the same term, 'deepfake'. AI generated, ready-to-use sound effects allow recorded audio in categories urbansounds, nature sounds and human sounds, are a major threat to democracy.

2. LITERATURE REVIEW

Political deepfake is the creation of false media contents (image, video, audio) for the purpose of misinforming the political audience, and gaining political advantage. Otherwise known as propaganda, deepfake has brought this techniqueclose to the people as well as made it more difficult to distinguish between real and fake media content. Some examples of deepfake deployment in politics are reviewed in this section. Detection methods proposed by different authors are alsoreviewed.

Deepfake use in Politics

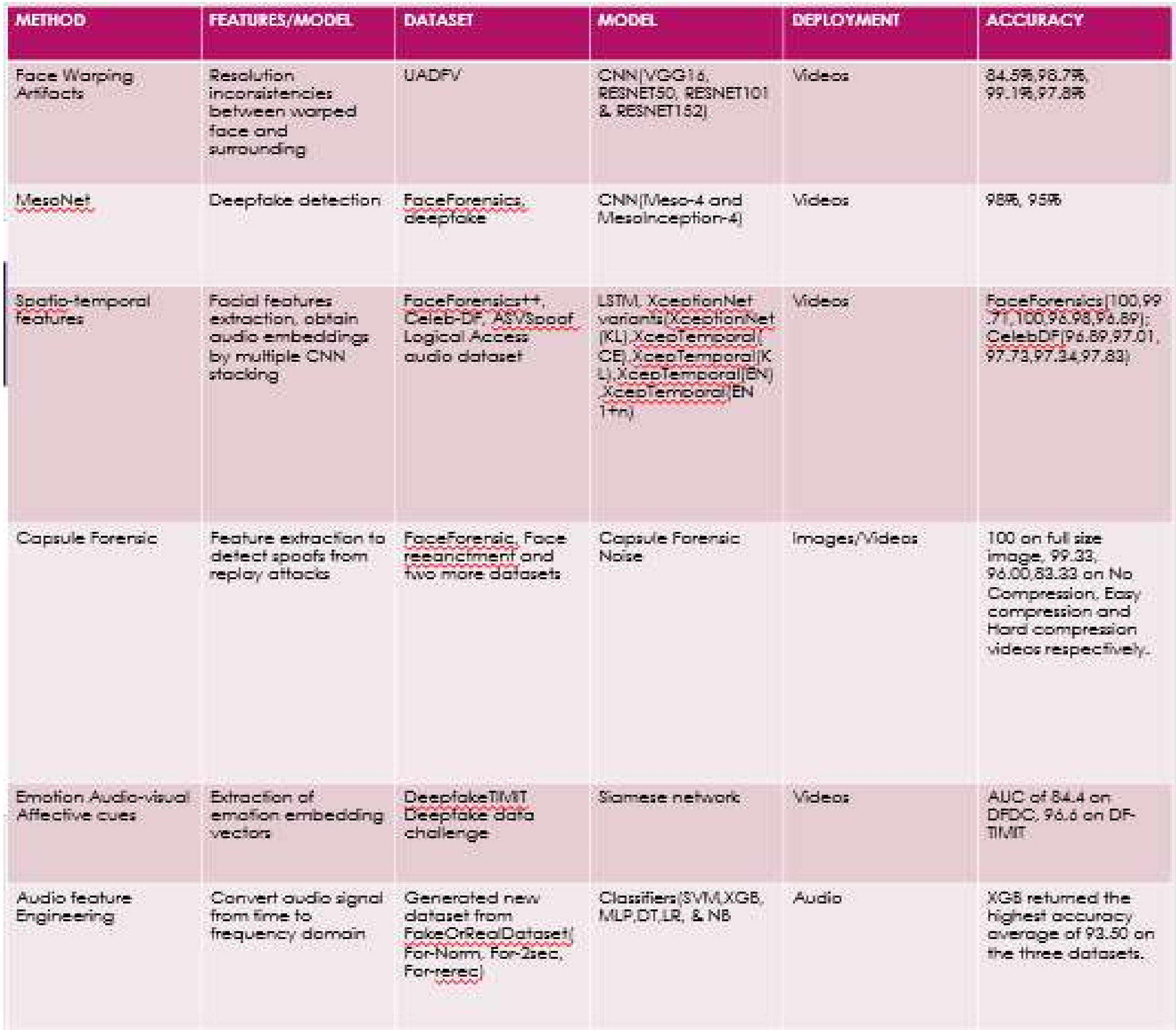
In 2018, a viral video showed former president Barak Obama calling the then incumbent president, Donald Trump 'dipshit' [4]. Also in 2018, a video of President Trump speaking about the Paris climate agreement was created and posted by Belgian political group, Socialistische Partij Anders (Sp.a), with the intention of misinforming the people and inciting them to sign a petition calling on the government to be more proactive in their action towards climate change. In the video, president

Trump was portrayed to say "we all know that climate change is fake" [5]. In the same year, a manipulated video of speaker Nancy Pelosi was shared by lawyer to President Trump, Rodulph Giuliani which portrayed the speaker as slurring in her speech. The fake video was accompaniedby a tweet "what is wrong with Nancy Pelosi, her speech is bizarre" [5]. PresidentTrump followed the misinformation up by sharing another slurred video of Pelosiin a 20-minute news conference with a tweet "PELOSI STAMMERS THROUGH NEWS CONFERENCE" [6] cited in [7]. According to [8], Volodymyr Zelensky, the president of Ukraine was seen in a video asking his troops to put down their weapons. This video was however found to be deepfakeas the Ukrainians quickly responded, but not before it had gone viral. Similarly, another deepfake video showed president Vladimir Putin of Russia declaring forpeace [9]. Also in 2021, a call from Leonid Volkov, chief of staff to imprisoned anti Putin politician, Alexei Navalny to European politicians was found to be a hoax carried out by an imposter named Alexei Stolyarov. [9].

3. Deepfake Creation

Like any new technology, experts in the field develop methods, lapses in introduced methods are discovered and improved upon. Hence deepfake tools have evolved over the years. A summary of some of the tools as presented in a survey by [10] is displayed in table 1 below: A summary of deepfake creation algorithms is presented in table 1 below:

Table 1: Some Deepfake creation tools Source: [10]



METHOD	FEATURES/MODEL	DATASET	MODEL	DEPLOYMENT	ACCURACY
Face Warping Artifacts	Resolution inconsistencies between warped face and surrounding	UADFV	CNN(VGG16, RESNET50, RESNET101 & RESNET152)	Videos	84.5%,98.7%, 99.1%,97.8%
MesoNet	Deepfake detection	FaceForensics, deepfake	CNN(Meso-4 and MesoInception-4)	Videos	98%, 95%
Spatio-temporal features	Facial features extraction, obtain audio embeddings by multiple CNN stacking	FaceForensics++, Celeb-DF, ASVSpoof Logical Access audio dataset	LSTM, XceptionNet variants(XceptionNet (KL),XcepTemporal(CE),XcepTemporal(K L),XcepTemporal(EN) ,XcepTemporal(EN 1+n)	Videos	FaceForensics(100,99 .71,100,96.98,96.89); CelebDF(96.89,97.01, 97.73,97.34,97.83)
Capsule Forensic	Feature extraction to detect spoofs from replay attacks	FaceForensic, Face reeanctment and two more datasets	Capsule Forensic Noise	Images/Videos	100 on full size image, 99.33, 96.00,83.33 on No Compression, Easy compression and Hard compression videos respectively.
Emotion Audio-visual Affective cues	Extraction of emotion embedding vectors	DeepfakeTIMIT Deepfake data challenge	Siamese network	Videos	AUC of 84.4 on DFDC, 96.6 on DF-TIMIT
Audio feature Engineering	Convert audio signal from time to frequency domain	Generated new dataset from FakeOrRealDataset(For-Norm, For-2sec, For-rerec)	Classifiers(SVM,XGB, MLP,DT,LR, & NB	Audio	XGB returned the highest accuracy average of 93.50 on the three datasets.

The paper captures major deep learning tools deployed in the creation of deepfakes such as autoencoder and decoder pairs used to achieve dimensionalityreduction and image compression, improvements using generative adversarial network (GAN) used to reduce artifacts visibility, as well as multitask convolutional neural network (CNN) used to improve face reliability and alignment.

4. DEEPFAKE DETECTION

A look at some methods through which deepfake has been detected in the past isrequired in order to provide a wholesome media literacy of this technology. Someof the proposed methods do not require expert technological skill but can still provide an escape route. [5] asserts that when you are having a video call and suspect the identity of the other person, ask them to turn to the side. This is because the software deployed in facial pose estimation does not have a perfect presentation at acute angle. [7] identified three sources of information through which deepfakes may be detected: context, audiovisual imperfections (technological glitches), and content. Contextual information such as during a war (president Putin and President Zelensky cases) calls for deep scrutiny. Audiovisual imperfections or artifacts are glitches such as found when video is viewed at acute angle. The third information the authors suggested is the contentof the media itself. Some deepfake videos include content that inform therecipient that the shared media is not original, or its intended nature.

However, this section may be edited out in multiple shared content thereby increasing misinformation. [11] investigated the presence of deepfakes in video contents bytesting known deepfaked video and the original versions on a deepfake model developed by one of the winners of Kaggle deepfake detection challenge. The model's output is classified as REAL if the score is close to 0 and FAKE if the score is close to 1. An application, fifty one was used to visualize the output. This model classified some videos correctly, but failed to detect deepfake in others. According to the author, research on models seeking to detect deepfakes is still at an early stage. He agrees with MIT researcher Matt Groh's suggestion on detecting deepfakes by particularly taking close look at the following:

Face: to observe glitches, eyebrows muscle movements, and facial muscle movements when they talk, as well as if the hair seem to fall naturally.

Audio: to observe if audio, facial expression and lips sync.Mouth: to observe unreal looking cuts and jumps.

Lighting: to observe portions of video where lighting does not match other scenes,are there reflections on illuminating surfaces such as glasses. [24]

Audio deepfakes are a major threat and other researchers at the University of Florida developed a technique that measures the acoustic and fluid dynamic differences between voice samples created organically by human speakers and those createdsynthetically by computers. More sophisticated deepfakes methods however, beatthe simple artifact-based detection techniques.

5. CONCLUSION

This is a situational awareness reporting of the deepfake technology and areas where it has been applied, including methods of creation and detection. The need to be aware of this technology is considered necessary for careful scrutinyof digital contents and easier detection of fake or disinforming content. The processes identified in research however, require some technical knowledge, hence the need for professional consultancy in cases of suspicious fake mediacontent.

EndNote/Copyright Status
This work was previously published in the JOURNAL OF SCIENTIFIC RESEARCH VOL.8, 2020

REFERENCES

1. S. S. Sansgiry, M. Bhosle, and K. Sail, "Factors that affect academic performance among pharmacy students," *Am. J. Pharm. Educ.*, vol. 70,no. 5, pp. 104–104, Oct. 2006, doi: 10.5688/aj7005104.
2. C. Audie, "Tracking Down Fake Videos : NPR," 2018. https://www.npr.org/2018/09/25/651560073/tracking-down-fake-videos (accessed Mar. 11, 2020).
3. K. Fagan, "Deepfake: Fake Obama Video Calling Trump Dipshit Is a Disturbing Trend," *businessinsider.com*, 2018. https://www.businessinsider.com/obama-deepfake-video-insulting-trump- 2018-4?r=US&IR=T
4. J. G. A. L. Foley, "The best deepfake examples | Creative Bloq," https://www.creativebloq.com/features/deepfake-examples
5. S. Mervosh, "Distorted Videos of Nancy Pelosi Spread on Facebook and Twitter, Helped by Trump - The New York Times," *The New York Times*, 2019. https://www.nytimes.com/2019/05/24/us/politics/pelosi-doctored- video.html (accessed Mar. 11, 2020
6. M. Appel and F. Prietzel, "The detection of political deepfakes," *J. Comput.Commun.*, vol. 22, no. 4, pp. 1–13, 2020.
7. R. Metz, "Facebook, YouTube, and Twitter remove Zelensky deepfake | CNNBusiness," *CNN Business*, 2020. https://edition.cnn.com/2020/03/16/tech/deepfake-zelensky-facebook- meta/index.html (accessed Mar. 16, 2020).
8. V. James, "'Deepfake' that supposedly fooled European politicians was just alook-alike, say pranksters - The Verge," *The verge*, 2020. https://www.theverge.com/2020/4/30/22407264/deepfake-european- polticians-leonid-volkov-vovan-lexus (accessed Mar. 11, 2020).
9. R. Redman, "The ethics of digital humans | Creative Bloq," 2017. https://www.creativebloq.com/news/the-ethics-of-digital-humans (accessed Mar. 16, 2019).
10. F. Iqbal, A. Abbasi, A. R. Javed, Z. Jalil, and J. Al-Karaki, "Deepfake AudioDetection via Feature Engineering and Machine Learning," *CEUR Workshop Proc.*, vol. 3318, 2020.
11. T. T. Nguyen *et al.*, "Deep Learning for Deepfakes Creation and Detection: ASurvey," Sep. 2019, doi: 10.1016/j.cviu.2022.103525.
12. E. Hofesman, "Have Deepfakes influenced the 2020 Election? | by Eric Hofesmann | Voxel51 | Medium," *Medium.com*, 2020. https://medium.com/voxel51/have-deepfakes-influenced-the-2020- election-c0fc890aca0f (accessed Mar. 12, 2020).
13. L. Blue and P. Traynor, "Audio deepfakes are a major threat — here's how researcherexpose them," *THE CONVERSATION*, 2020. https://www.inverse.com/innovation/deepfake-audio-detection (accessed Mar. 16, 2020).