DeepVoice & DeepFakes: Exploiting AI-Generated Audio & Video – Boon & Bane

Ms. K. Visalini
Researcher, International Research Center
Kalasalingam Academy of Research & Education, Virudhunagar, Tamil Nadu
info@kvisalini.com

Introduction:

We all are aware of the fast-pace that AI is exploiting various domains. All of those can never be enumerated in a single article yet; the most trending advancement is discussed in this article in a bipolar view point. Ardent followers of Artificial Intelligence & Machine Learning will be well aware of the raising upswing in the technology in the early 2018 – AI generated audio and video. It enthrals many at the same time, intimidates most. Let us discuss this in both the viewpoints.

It was bone-chilling shock catered by the Chinese technology giant Baidu in mid-March 2018. They released a white-paper in Arxiv[1], where they claimed that they can clone the voice of any Homo-Sapien with a training audio of less than five seconds, with implications for both trust and security. According to the article published, the AI algorithm developed named "Deep Voice" was capable of cloning the voice of any individual to believable extent such that, many human subjects as a part of study experiment conducted during the beta test considered it truthful.

Deep Voice - Evolution:

Almost a year ago, the system required about half an hour of audio input to generate, a fake audio clip. But, now after much effort of the developers, the technology is now able to create much better results with training input of just 3.7 seconds that accounts for shorter than a whole sentence, especially just one sample to learn from.

Baidu did not stop this interesting evil inside their labs; they uploaded an array of code and system demos into Github[2], demonstrating the capabilities of Deep Voice (the sample demos, range from hilarious to creepy). These include voice cloning and various manipulations, such as changing the voice from male to female or even British to American.

Soon, it stimulated many artistic Machine Learning enthusiasts to fork on their versions. Deep Speech[3], that asks to you to speak less and makes you speechless. It was a work of Mozilla, of about 1400 commits and 63 contributors in GitHub, based on Baidu's work.

One of the most notable work, is of Lyrebird.ai, a Montreal-based start-up that kick-started in 2017, with a team of four individuals (including an Indian guy), in a small rental room in the other side of a coffee shop will make you stunned with their artefact. The growing start-up asks the user to under a three step process: First, the user is required to provide a voice clip for just a minute; then, processes it in the next minute; finally, the user can type all that he wants the AI to speak. Their awesomeness upraised them to feature in Bloomberg [4].

Negative Outlook:
While Baidu writes that 'Voice cloning is expected to have significant applications in the direction of personalisation in human-machine interfaces', but there arises natural concerns about identity theft and privacy.

Tom Harwood, CPO and co-founder at voice security solutions provider Aeriandi, said: "This technology is poised to transform personalisation in human-machine interfaces, but it raises serious concerns about voice biometric security systems. Soon, criminals will need just a few seconds of someone's voice to cheat a voice recognition security system - voice biometric authentication will be rendered useless. Organisations need to be thinking now about how they can implement new technologies to ensure they stay ahead of the curve. Voice fraud detection technology is the primary candidate, as it looks at far more than the user's voice print; it considers hundreds, if not thousands, of other parameters. For example, is the phone number being used legitimate? Has it been used fraudulently before? Increasingly, phone fraud attacks come from overseas. Voice fraud technology has been proven to protect against this as well as domestic threats."

Positive Usages:
But where can we implement Deep Voice, where it results in no harm to either the society or an individual? It can be used to create the best experience of late actors, in case if we want them to play any role in future movies. As, we know, their visual imagery can be generate with motion capture, the AI concocted voice of the dead actor can provide the most realistic experience to the audience, if and only if, we accept that it hits badly at the daily bread of mimicry and voice actors.

DeepFakes – Real fake face of AI:

The development of Deep Voice to require only minimal training speech could further raise distrust in internet media - mimicking the 'deepfakes' fake celebrity porn videos that began popping up earlier this year. Moreover, earlier in April

2018, Barack Obama called President Trump a “total and complete dipshit?”. Okay, that was an AI generated video, that exacted the former President of the United States. It was enacted by Jordan Peele, with BuzzFeed as a warning to us all about how deep fakes could be used to distort reality, and it has been a constant topic around political and societal stability, how we can all protect ourselves against being duped, and the potential consequences if we can’t.[5]

Fake videos can now be created using a machine learning technique called a “generative adversarial network”, or a GAN. A graduate student, Ian Goodfellow, invented GANs in 2014 as a way to algorithmically generate new types of data out of existing data sets. For instance, a GAN can look at thousands of photos of Barack Obama, and then produce a new photo that approximates those photos without being an exact copy of any one of them, as if it has come up with an entirely new portrait of the former president not yet taken.

The use of this machine learning technique was mostly limited to the AI research community until late 2017, when a Reddit user who went by the moniker “Deepfakes” – a portmanteau of “deep learning” and “fake” – started posting digitally altered pornographic videos. He was building GANs using TensorFlow, Google’s free open source machine learning software, to superimpose celebrities’ faces on the bodies of women in pornographic movies.

A number of media outlets reported on the porn videos, which became known as “deep fakes”. In response, Reddit banned them for violating the site’s content policy against involuntary pornography. By this stage, however, the creator of the videos had released FakeApp, an easy-to-use platform for making forged media. The free software effectively democratized the power of GANs. Suddenly, anyone with access to the internet and pictures of a person’s face could generate their own deep fake.

Negative Concerns:
“The marketplace of ideas already suffers from truth decay as our networked information environment interacts in toxic ways with our cognitive biases,” a report reads. “Deep fakes will exacerbate this problem significantly.”

In August, an international team of researchers affiliated with Germany’s Max Planck Institute for Informatics unveiled a technique for producing what they called “deep video portraits”, a sort of facial ventriloquism, where one person can take control of another person’s face and make it say or do things at will. A video accompanying the research paper depicted a researcher opening his mouth and a corresponding moving image of Barack Obama opening his mouth; the researcher then moves his head to the side, and so does synthetic Obama.

Positive Perspective:
Christian Theobalt, a researcher involved in the study, told that he imagines deep video portraits will be used most effectively for accurate dubbing in foreign films, advanced face editing techniques for post-production in film, and special effects. In a press release that accompanied the original paper, the researchers acknowledged potential misuse of their technology, but emphasized how their approach – capable of synthesizing faces that look “nearly indistinguishable from ground truth” – could make “a real difference to the visual entertainment industry”.

Hany Farid, professor of computer science at the University of California, Berkeley, believes that although the machine learning-powered breakthroughs in computer graphics are impressive, researchers should be more cognizant of the broader social and political ramifications of what they’re creating. “The special effects community will love these new technologies,” Farid told[6].

New Problem to face for Cyber Security Specialists:
Farid, who has spent the past 20 years developing forensic technology to identify digital forgeries, is currently working on new detection methods to counteract the spread of deep fakes. One of Farid’s recent breakthroughs has been focusing on subtle changes of color that occur in the face as blood is pumped in and out. The signal is so minute that the machine learning software is unable to pick it up – at least for now.
Isn’t this new opening, in cyber security a positive perspective? It really is, for passionate cyber security enthusiasts.

Conclusion:

It is not the first time, the world encounters forgery, it has been there for eras. Every technology has its own pros and cons. Though in this concern, the negativity outweighs the positive aspects, if people are enlightened and if proper solutions are attained such that, if another AI could possibly differentiate a legitimate video and fake one, then that should be called a true development.

References:

[1] Arik, S.O., Chrzanowski, M., Coates, A., Diamos, G., Gibiansky, A., Kang, Y., Li, X., Miller, J., Ng, A., Raiman, J. and Sengupta, S., 2017. Deep voice: Real-time neural text-to-speech. arXiv preprint arXiv:1702.07825.
[2] GitHub page of Baidu-Research – Deep Voice; https://github.com/baidu-research/deep-voice

[3] GitHub page of Mozilla – DeepSpeech; https://github.com/mozilla/DeepSpeech
[4] This AI Can Clone Any Voice - Including Yours, Hello World – Season 1 Episode 15. https://www.youtube.com/watch?v=VnFC-s2nOtI
[5] This new deep fake video is both advertising and a piece of art – Fast Company https://www.fastcompany.com/90279597/this-new-deep-fake-video-is-both-advertising-and-a-piece-of-art
[6] You thought fake news was bad? Deep fakes are where truth goes to die – The Guardian https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth

About the Author:

Ms. K. Visalini is a five World Records and 13 International Certifications holder. She was directly admitted in the B.Tech program as an Exceptional Candidate, after her 8th grade at school. She had completed the four year B.Tech degree programme in three years with a GPA of 9.6 out of 10 and excelled with Honours.

The Indian Prime Minister Shri. Narendra Modi lauded her with the quote "Visalini, Whatever you have achieved in this young age is a great service to our country India" and wished her luck for her future endeavours. She was honoured twice at the age of 3 and 14, for her achievements by the former President of India, Dr. APJ. Abdul Kalam.

At the age of 15, Visalini was invited by the Director – ISRO (Indian Space Research Organization), to deliver a technical lecture to 700+ ISRO Scientists on "Big Data & AI for ISRO". Her lecture received huge applause & standing ovation from the scientists and a model of the Mangalyaan Satellite (Sent to MARS by ISRO) was presented to her. She was entrusted with a vital project for ISRO: ISAC-VNMS (ISRO Satellite Center – Visalini's Network Management System) and was provided two years duration, but she had completed within 35 days & dedicated to the Nation at the age of 15.

From the age of 11, Ms. K. Visalini was invited as a guest/invited speaker to deliver keynote addresses in 12 International Conferences, including two TEDx & Google India Summit. She is well described as an Enthusiastic & Passionate Researcher, working in Artificial Intelligence, Cognitive Neuroscience, Virtual Reality & Autonomous Navigation Agents. She is also working in multiple projects simultaneously, of which the most vital are: Emergency Rescue Equipment for Indian Soldiers, Rescue Equipment for Fishermen during Natural Calamities, SOS Equipment for Women, Assistive Technology for Mentally-Challenged and Visually-Impaired Students.

For more info pl. visit www.kvisalini.com