THE WEAPONIZATION OF DEEP FAKES: THREATS AND RESPONSES

Date: August 19, 2021

Disclaimer: This briefing note contains the encapsulation of views presented by the speaker and does not exclusively represent the views of the Canadian Association for Security and Intelligence Studies.

KEY EVENTS

On August 19, 2021, the Canadian Association for Security and Intelligence Studies (CASIS) Vancouver hosted a digital roundtable titled *The Weaponization of Deep Fakes: Threats and Responses* conducted by our guest speaker, Dr. Hany Farid, a professor at the University of California, Berkeley School of Information. The presentation was followed by a question and answer period with questions from the audience and CASIS Vancouver executives.

NATURE OF DISCUSSION

Presentation

Since the start of the COVID-19 pandemic, misinformation and conspiracies about the deadly COVID-19 disease quickly overtook the internet as swarms of individuals claimed a variety of different theories to be true, including microchipping vaccines and the governmental hoax of the disease. Although these nonsensical beliefs seem harmless, Dr. Farid outlined the true dangers and importance of these misconceptions with real-world data samples. Dr. Farid, then, described the process of creating deep fakes, otherwise known as synthetic media, of realistic appearing humans through artificial intelligence (AI) software, as well as why this is a cause for concern. The presentation concluded with Dr. Farid explaining the weaponization of deep fakes, the innovation of deep fake detecting techniques, the future challenges, and the steps global society can take to control these arising problems.

Question and Answer Period

The question and answer period was mainly focused on governmental and individual protective measures against deep fakes. However, Dr. Farid also

discussed the ethical application of weaponizing deep fakes, synthetic hair simulation, and regulatory surveillance in respect to aggressive socialism.

BACKGROUND

Presentation

During the span of the ongoing COVID-19 pandemic, government conspiracy theories and disease misinformation became increasingly popular with the introduction of COVID-19 vaccines in late 2020 and early 2021. However, this notion of "fake information" is not new to us. According to a 2015 conspiracy poll by Public Policy Polling, 50% of United States' citizens believe in at least one conspiracy theory. Examples of these theories include 4% of US citizens believing that Lizard people control our society, 15% feel that the government adds mind controlling technology into televisions, 19% think that the 9/11 US attack was staged by the government, and 37% assume that one of the world's growing issues, global warming, is a hoax. Despite easily dismissing these thoughts as irrational, they pose great dangers to our society because they destroy the public's trust in governments and scientific institutions, especially in times of crisis such as the current pandemic.

When citizens fail to trust the government, institutions, and academics, they are less likely to follow the guidelines provided by these organizations and institutions such as wearing masks or getting vaccinated. In a 2021 global survey conducted by Nightingale and Farid (2021), an estimated 12-18% of the 2,708 participants worldwide believed in falsified COVID-19 statements. More specifically, 18% of these participants were convinced that COVID stood for Chinese Originated Viral Infectious Disease which is one of the leading causes of the increase of violence against individuals of Asian descent (Nightingale & Farid, 2021). Additionally, 21% of survey participants believed that large doses of vitamin C protect against COVID-19 while 22% think that Bill Gates is utilizing vaccines to implant tracking microchips into people; these beliefs are used as reasons to not get vaccinated, potentially contributing to the continuation of COVID-19 deaths and a prolonged pandemic (Nightingale & Farid, 2021).

Currently, one of the leading sources of misinformation is social media. Individuals who rely on social media as their main source of news are 1.4 times more likely to trust falsified information (Nightingale & Farid, 2021). This is becoming progressively more concerning because the new and innovative wave of artificial intelligence is making the creation of synthetic media, also known as "deep fakes", more lifelike and easily accessible to the public. Through websites

such as thispersondoesnotexist.com, users can utilize the AI's programming of its generative adversarial network to create computer generated pictures of non-existent beings at the click of a button.

The generative adversarial network operates by pitting two of its systems, the generator and the discriminator, against each other with the common goal of creating a realistic-looking person. The generator's role in the network is to generate an image of a human from random pixels which will then be sent to the discriminator. It will then try to distinguish the differences between the generated image and real-world images taken from the internet. This cycle is repeated until the generator produces an image that the discriminator can no longer distinguish from the real images. A similar process is utilized to change a person's identity through face swap on video recordings. Here, the user will switch the individual's face (eyebrow to chin) in the video with the face of another whilst keeping everything else in the frame the same.

Because generation and modification software are so readily accessible online, it can be easily weaponized and used towards unknowingly vulnerable individuals. Non-consensual intimate pornography of women, where a women's likeness and features are edited onto sexually explicit content and then distributed, is just one instance of how this AI software can be used as a dangerous weapon. In a political context, deep fakes can be used to spread falsified news by impersonating a person of power such as political figures and promoting misinformation campaigns on social media. In addition, this deep fake technology can also cast doubt on evidence collected by law enforcement agencies as individuals may no longer trust video and photo evidence.

To counter these threats, academics like Dr. Farid have been exploring and developing software that utilizes behavioral mannerisms to distinguish between fake and real media. The employment of soft biometrics compares the movement of one's mouth with the rotation of their head from a real video, and then uses it to analyze the possibly fake video. Since the mouth is derived from a different video clip, but the head remains untouched, there will be deviations in the soft biometric measurements, which we can use to determine whether the video is fake or not. Phonemes, the fundamental mouth shapes used to create sounds, is another way analysts can use to differentiate the real from the fake. This is one aspect that synthetic media tends to miscalculate, and because it is such a small detail, it can be easily overlooked by the human eye. In addition, the movement of one's ear in comparison to their mouth can also be used to distinguish whether a video is fake or real.

Although software is being constructed to authenticate video footage, there are still many challenges that lie ahead and we, as a society, need to work together to prevent the weaponization of deep fakes. For now, behavioral mannerisms are an effective way to distinguish between real and fake footage. However, AI is a rapid and ever-growing field, and the new wave of technology is only going to get better at creating deep fakes. With the wide usage of social media world-wide, these fakes can reach millions in a matter of seconds through shares, likes, and comments. To counter this, better technology must be developed to combat these deep fakes; therefore, a natural arms race between synthetic media intelligence and distinguishing systems has begun. In addition to creating better tools to counter deep fake media, corporations need to be held responsible for what is circulating on their networks, and governmental regulation of these sites need to be developed and enforced in order to limit the weaponization of deep fakes. In terms of the next generation, tools need to be formulated to promote internet safety and used to educate individuals before it's too late.

Question and Answer Period

The question and answer period began with a question concerning the ethical application of deep fake technology weaponization. In response, Dr. Farid explained that deep fake technology is not inherently bad; however, there are people in the world that will utilize this technology in harmful ways. Therefore, the responsibility falls onto the creators to consider the potential risks that their system poses and innovate guidelines to mitigate them before releasing the product to the public.

The discussion then proceeded to a conversation about individual and governmental protective measures that can be employed to prevent and counter victimization from deep fakes. In respect to the upcoming wave of technology, Dr. Farid acknowledged that it is impossible to predict what will come next; however, we must be proactive and start utilizing safeguards, such as emphasizing safety protocols during its creation, in order to protect vulnerable individuals from the new age of technology. In terms of individual protective measures, Dr. Farid stated that it is too late for this generation to combat deep fake technology because AI has already evolved to the point where a single image of an individual can be used to create deep fake videos of them through a software called Puppetmaster. Nonetheless, not all hope is lost for the next generation if we start enforcing social media regulations, making corporations responsible for the online content they circulate, and educating the next wave of technology users on internet safety.

As for governmental regulation of deep fakes, Dr. Farid suggested that officials tend to focus less on the concept of deep fakes and more on the underlying issue of misinformation that it causes. Although laws are slowly being passed to regulate deep fake media and discussions of internet regulation are circulating at the federal level, the bills passed are faulty and provide no coherent federal response. In order to adequately manage the internet, the government must find the right balance between keeping the internet open and free whilst regulating the online world and keeping it safe. This may be tricky, but Dr. Farid noted that the biggest problem we face when it comes to deep fake regulation discussion is the disagreement of the problem's root cause. In terms of the private sector, technology is being developed at these social media corporations, and they are created half-heartedly and with no willingness to make a change.

With respect to law enforcement and evidence tampering, Dr. Farid expressed that law enforcement agencies may never develop a reliable way to fully authenticate digital evidence, but to work around this, software can be used to authenticate real footage by using control capture technology. With control capture technology, the video footage and all its pixels are cryptically signed with the date, time, and geo-tag during the time of the recording. By doing this, law enforcement will no longer have the responsibility of authenticating the footage after the time of recording because the camera would have already authenticated the footage at the time of recording. Dr. Farid suggested that police body cams and CCTVs should utilize this software to avoid doubt about the reliability and authenticity of these footages when being analyzed.

The last question tackled the concept of aggressive socialism regarding the regulation surveillance that Dr. Farid is continuously promoting. Dr. Farid noted that we are already subjected to constant surveillance. Surveillance capitalism is an enticing feature of today's business model as social media content and networks are advertised as "free", but in reality, we are giving away our privacy in exchange for this software. Therefore, Dr. Farid stressed that paying for media networks is recommended as they would be more likely to respect privacy, and he hopes that in the coming decades, a different business model is developed that emphasizes better corporate leadership.

KEY POINTS OF DISCUSSION

Presentation

- The growing problem of conspiracy theories and misinformation is not the fact that people believe in them, but the real danger stems from the erosion of governmental and institutional trust.
- Synthetic media technology utilizes generative adversarial networks, a cycle of generating and distinguishing photos or videos until they satisfy the authenticity threshold, in order to modify video footage and create photos of realistic but non-existent beings.
- Weaponization of deep fakes through AI includes non-consensual intimate pornography of women, evidence tampering, and misinformation campaign, thus causing a new age arms war between weaponizors and protectors.
- Behavioural mannerisms such as head rotations, lip movements, ear motions, and phonemes can be measured and utilized to distinguish fake deeps from real media.
- With the easy accessibility of these deep fake software and wide reach of social media networks, there are many dangers that lie ahead in the future; however, through regulation, education, and safety technology, these threats may be limited for the future generation.

Question and Answer Period

- Responsibility must fall on the software inventors to adequately assess the dangers of their programs and create protocols to mitigate risks before releasing the technology out to the public as this software may be, unknowingly, used as weapons.
- It's too late to enact protective measures for this generation of tech users because software such as PuppetMaster only requires a single photo of the individual to create deep fake videos.
- Governments must focus more on the concept of misinformation and less on deep fakes; the underlying problem is what must be tackled in order to properly regulate the internet.
- Control capture technology is the best way to combat evidence tampering as it specially signs each pixel with the date, time, and geo location of the time of recording, allowing law enforcement agencies to easily authenticate and analyze the footage.
- Surveillance capitalism, a feature in our current business model, is widely accepted in exchange for free media content; however, the hefty price of

privacy is not worth the media content we get in return. Therefore, the hope is that in the coming decades, a new business model will be developed that will hold corporations liable for their networks.

References

Nightingale, S., & Farid, H. (2021, January 27). *Examining the Global Spread of Covid-19 Misinformation.* Cornell University. https://arxiv.org/abs/2006.08830v2

Published by the Journal of Intelligence, Conflict, and Warfare and Simon Fraser University

Available from: https://jicw.org/