The Rising Tide

Surfacing new and emerging crime trends and deviant behaviours within and beyond our shores

By Crime, Investigation and Forensic Psychology Branch, Home Team Behavioural Sciences Centre

Issue #3 2019

Deep Fakes: A New Frontier for Falsified Digital Media

Rana Ayyub, an outspoken investigative journalist in India, had tolerated her fair share of online hatred and misinformation that were deliberately targeted at her. Yet, everything changed in April 2018 when a pornographic video of her began to circulate across the country. It was in fact a fabricated video of a pornographic actress with Rana's face digitally stitched over hers. Yet, the harm she sustained was real – her reputation was tarnished, she was hospitalised for an acute stress reaction, and she was left with a deep psychological scar because of one 'deep fake' video that went viral (Ayyub, 2018).

The creation and spread of online falsehood is not new in today's digital landscape. Yet, advances in artificial intelligence and machine learning have resulted in the emergence of 'deep fakes' – a new generation of digitally falsified media that has further thinned the line between objective reality and a digitally-created one.

About Deep Fakes

Deep fakes refer to hyper-realistic, digitally altered videos and audio that are generated by deep learning software. Deep fake technology works through machine-learning algorithms that recognise faces and voices from multiple input data and maps them over existing video and audio recordings. The output is a life-like, digital creation that makes a person appear to say or do something that he or she has never done in real life (Chesney & Citron, 2018).

This technology has existed and has been applied in the AI research community and film-making industry for years. It was only until late 2017 that a Reddit user known as "Deepfakes" released a user-friendly software that democratised the creation of deep fakes among the internet community (Schwartz, 2018). Adopters have predominantly used it to digitally insert faces of public figures or Hollywood personalities into movie scenes or superimpose faces of celebrities onto pornographic films for sharing within online communities. These are typical targets as videos of them are abundant on the web and this myriad of data increases the accuracy of the software to generate realistic deep fakes.

Filmmaker Jordan Peele and Buzzfeed created a realistic digital impersonation of former President Barrack Obama warning about fake news and the need for vigilance (Mack, 2018).

Deep fake technology is becoming increasingly accessible to any layperson. Professional software such as FakeApp and DeepFaceLab are easily obtainable on the web, while video tutorials offer both basic and advanced guides on making high quality deep fakes (Derpfake, 2019).

Potential Harms of Deep Fakes

Such a tool that has the capability to create false realities would harbour many potential misuses. Chesney & Citron (2018) have listed the following as potential causes for concern:

Harm to Individuals

- **Exploitation and extortion** of money, confidential information, or private images and videos.
- **Humiliation and degradation** of victims leading to lasting psychological harm.
- **Sabotage of reputation** and future prospect.

Harm to Society

- **Eroding trust in public and private institutions** through attacks on credibility and reputation.
- **Inciting racial or religious tensions** or violence by targeting societal fault lines.
- **Causing public alarm and panic** that undermine safety and security.
- **Disrupting salient matters** concerning diplomacy and national security.

With the above dangers in mind, researchers are posed with the challenge to produce forensic detection methods to identify deep fake videos from actual footages. Yet, there are concerns that this will be an uphill battle as machine learning techniques can circumvent them at an equal if not faster rate. (Schwartz, 2018). Furthermore, deep fake technology continues to refine itself and innovate with advances in machine learning. For example, the algorithm to create full-body deep fakes that render realistic bodily motions have surfaced, though it is still at an early stage of development (Robitzski, 2018).

Implications for the HT

The ongoing fine-tuning and application of deep fake technology could become a future concern to Singapore if it is employed in the next generation of cybercrimes. Phishing or phony investment scams that impersonate public figure or notable personalities may become more convincing to the untrained eye with realistic impersonations endorsing these false products. Extortion scams may rely on digitally-altered videos to prey on the modesty and reputation of innocent victims for financial exploitation or data theft. Its application to revenge pornography is also of no exception as there are presently many deep fakes of doctored sex videos on the web.

The potential threat of deep fake videos reinforces the recurring need for sound measures to regulate and respond to the creation and spread of deliberate online falsehood. This could be in the form of increased collaboration with stakeholders like technology companies or digital forensic researchers to discourage, detect or remove digitally falsified content that can cause harm to individuals, institutions, or society. Furthermore, building a media-literate citizenry will continue to play a key role in equipping the public with the knowledge and skills to discern between real and deep-fake media and to know how to respond to them.

References

Ayyub, R. (2018, November 21). I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me. *Huffington Post UK*. Retrieved from https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316

Chesney, R., & Citron, D. K. (2018). "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *107 California Law Review (2019, Forthcoming); U of Texas Law, Public Law Research Paper No. 692; U of Maryland Legal Studies Research Paper No. 2018-21*. Available at SSRN: https://ssrn.com/abstract=3213954

Derpfake (2019, January 14). *Deepfakes Tutorial | Episode One | Overview* [Video file]. Retrieved from https://www.youtube.com/watch?v=cVcyghhmQSA

Mack, D. (2018, April 17). This PSA About Fake News From Barack Obama Is Not What It Appears. *Buzzfeed News*. Retrieved from https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed

Robitzski, D. (2018, December 12). These Full-Body Deepfakes are Like Nothing We've Ever Seen. *Futurism*. Retrieved from https://futurism.com/full-body-deepfakes

Schwartz, O. (2018, November 12). You thought fake news was bad? Deep fakes are where truth goes to die. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth

Image provided by Pixabay.com