NATIONAL CAPITAL REGION
THREAT INTELLIGENCE CONSORTIUM

Intelligence Bulletin

August 29, 2019

Product No. 2019-02-022
NTIC SIN No. 2.7 | HSEC No.7.3, 7.4, 10.8

Detecting Deep Fakes

As research institutions and government agencies attempt to detect deep fakes—manipulated videos containing false information—producers are likely to counter the advances and produce more sophisticated deep fakes. Deep fakes use still images of an individual and apply machine learning algorithms to synthesize facial movements into video form. Currently, video images lacking natural human movements offer clues to detect a deep fake; however, rapidly evolving technology may soon erase these indicators.

- In June 2018, researchers at the State University of New York applied these movement clues and artificial intelligence (AI) techniques to track eye blinking in videos, achieving a 95 percent detection rate for deep fake videos. However, shortly after this method was identified, the synthesis techniques used to develop deep fakes were altered to eliminate this flaw.

- Researchers at the University of California at Berkley and Southern California built an AI-system that detects deep fakes using biometric models to determine if "real" facial and head movements have been altered. The developers openly acknowledge that deep fake creators likely will eventually adapt to this form of detection, but have decided not to release the code behind the detection method to slow down discoverability.

Government Efforts

The US Defense Advanced Research Project (DARPA) established the Media Forensics program to develop tools to detect deep fakes. DARPA is bringing researchers together and providing funding to mitigate the risk posed by deep fakes. Experts are examining heat mapping and light levels in altered videos and searching for the absence of physiological actions, such as blinking and breathing, with the goal of improving future detection and mitigation strategies for deep fake videos.

In 2019, DARPA announced the launch of a second program, Semantic Forensics, designed to spot errors automated systems used to manipulate media make when processing large amounts of data, such as mismatched earrings.

It is still possible to identify less sophisticated and poorly made deep fakes as these videos use weak configurations of algorithms or a small number of photographs that reveal easily identifiable flaws. Flickering facial areas, blurred facial or body features, and boxes around the face may be the clues a viewer needs to spot a deep fake.

Blurred facial areas in poorly made deep fake (Source: YouTube)

This is the NTIC's third product in a series focused on deep fakes at the UNCLASSIFIED level. The first product provided an overview of deep fake technology and the second focused on how it can be used to spread disinformation. Check out this publication and more at ncrintel.org.