FORTUNE

How Faking Videos Became Easy — And Why That's So Scary

By **BLOOMBERG** September 11, 2018

A minute-long video of Barack Obama has been seen more than 4.8 million times since April. It shows the former U.S. president seated, with the American flag in the background, speaking directly to the viewer and using an obscenity to refer to his successor, Donald Trump. Or rather, his lips move as the words are spoken. The video is actually a so-called deep fake made by actor-director Jordan Peele, who impersonated Obama's voice. Peele created the video to illustrate the dangers of fabricated audio and video content depicting people saying or doing things they never actually said or did. Researchers at New York University describe deep fakes as a "menace on the horizon."

1. What constitutes a deep fake?

While manipulation of digital files is nothing new, this breed of believable fakery is accomplished using computer programs that employ a form of artificial intelligence. An algorithm is trained to recognize patterns in actual audio or visual recordings of a particular person, a process known as deep learning. As with doctored images, a piece of content can be altered by swapping in a new element — such as someone else's face or voice — and seamlessly joining the two. The manipulations are most misleading when combined with voice-cloning technology, which breaks down an audio recording into half-syllable chunks that can be reassembled into new words — the same method that's used to create voice assistants like Apple's Siri and Amazon's Alexa.

2. Why are some fakes more believable than others?

The bigger the library of content a deep-learning algorithm is fed with, the more realistic the phony can be. The Obama deep fake required 56 hours of sample recordings. Apple recorded 10 to 20 hours of speech to create Siri. According to a number of reports, voice clones can be made from little more than a few seconds of material.

3. How did this technology spread?

Motherboard, a Vice publication, reported in December that a Reddit user called "deepfakes" had made publicly available an algorithm for making fake videos using open-source code. Previously, the technology was the domain of academics and researchers, but now anyone could use it. It took off as a means to create phony pornography, usually with the faces of female celebrities mapped on to porn stars' bodies to depict sex acts that never took place. Reddit banned the user "deepfakes," but the technology spread and is now readily available on apps such as FakeApp.

4. Apart from pornographers, who's making fakes?

In what's been called a technological arms race, universities and research companies are developing the technology to test the power of deep fakes and to beat nefarious practitioners to it. Researchers at Carnegie Mellon University recently created a system that can transfer characteristics, such as facial expressions, from a video of one person to a synthesized image of another. China's Baidu and a handful of startups including Lyrebird and iSpeech have been selling voice cloning for commercial use in human-machine interfaces.

5. Why are people worried?

The fear is that deep fakes could unduly destroy reputations and even set off unrest. Imagine falsified videos depicting a presidential candidate molesting children, a police chief inciting violence against a minority group, or soldiers committing war crimes. High-profile individuals such as politicians and

business leaders are especially at risk, given how many recordings of them are in the public domain. For ordinary people, especially women, the technology makes revenge porn a possibility even if no actual naked photo or video exists. Once a video goes viral on the internet, it's almost impossible to contain. An additional concern is that spreading awareness about deep fakes will make it easier for people who truly are caught on tape doing or saying objectionable things to claim that the evidence against them is bogus.

6. Can the fakes be detected?

The kind of machine learning that produces deep fakes can't easily be reversed to detect them. Researchers have identified clues that might indicate a video is inauthentic — if the speaker hasn't blinked for a while or seems slightly jerky, for example — but such details could easily slip a viewer's notice. By enhancing the color saturation on a video of a person, it's possible to detect his or her pulse from the almost invisible change in facial skin; an image made from a mishmash of clips would have an irregular or non-existent blood flow. The U.S. Defense Department is developing tools to counter deep fakes.

7. Are there benevolent uses?

Yes. Scottish firm CereProc creates digital voices for people who lose their own through disease, and vocal cloning could serve an educational purpose by recreating the sound of historical figures. A project at North Carolina State University synthesized one of Martin Luther King Jr.'s unrecorded speeches. CereProc created a version of the last address written by President John F. Kennedy, who was assassinated before delivering it. The John F. Kennedy Library rejected the recording, though, saying it didn't meet its guidelines.