Deepfakes: Fighting fake videos, from Silicon Valley to Washington

January 3 2019, by Levi Sumagaysay

Credit: CC0 Public Domain

Whether it's a video showing someone else's face on a different body or former President Barack Obama saying things he didn't really say, "deepfakes" are now a thing.

Another popular use of such manipulated videos: fake porn, featuring everyone from actress Scarlett Johansson to possibly anyone whose

photos are available online.

Johansson, who has unwittingly starred in a supposedly leaked pornographic video that the *Washington Post* says has been viewed more than 1.5 million times on a major porn site, feels the situation is hopeless.

"Nothing can stop someone from cutting and pasting my image or anyone else's onto a different body and making it look as eerily realistic as desired," she told the *Post* in a story published a few days ago. "The Internet is a vast wormhole of darkness that eats itself."

That's why researchers are hard at work trying to figure out ways to detect such fakery, including at SRI International in Menlo Park.

Manipulating videos used to be the stuff of movies, done by Disney, Pixar and others, said Bob Bolles, program director at the AI center's perception program at SRI. "Now face-swap apps are getting better and better."

Bolles recently showed off how he and his team are trying to detect when video has been tampered with.

The team looks for inconsistencies between video and the audio track—for example, watching whether a person's lip movements match the sounds of the video. Think videos that aren't as obvious as those old, badly dubbed kung fu movies.

The team also tries to detect sounds that may not jibe with the background, said Aaron Lawson, assistant director of the speech lab at SRI. For example, the video may show a desert scene but reverberations can indicate the sound was recorded indoors.

"We need a suite of tests" to keep up with hackers who are bound to keep figuring out new ways to keep fooling people, Bolles said.

The SRI team is just one of many among the organizations, universities and companies working to try to detect and even trace deepfakes under a DARPA program called MediFor, which is short for media forensics. Matt Turek, program director for MediFor at the Defense Advanced Research Projects Agency in Washington, D.C., said 14 prime contractors are working on the project, which was started in summer 2016. Those contractors have subcontractors working on MediFor, too.

"There's been a significant change in the last year or so as automated manipulations have become more convincing," Turek said. Artificial intelligence and machine learning have helped make tools to create deepfakes more powerful, and researchers are using AI to try to fight back.

"The challenge is to create algorithms to keep up and stay ahead of the technology that's out there," Turek said.

DARPA hopes the fruits of the MediFor program will be distributed far and wide, picked up by tech companies such as Facebook and YouTube, which handle a big fraction of the world's user-generated videos.

For example, a video published to YouTube by BuzzFeed this past April, featuring comedian and actor Jordan Peele ventriloquizing Obama—with the former president seemingly calling President Donald Trump an expletive—has more than 5.3 million views.

Turek ticked off other possible uses for deepfakes, such as to misrepresent products being sold online, or to fake out insurance companies over car accidents. He said there's been evidence of researchers using manipulated images so their scientific findings can get

published. And he predicts that eventually, tools could get so good that a series of deepfakes will be able to convince people of significant events that didn't happen—cue the conspiracy theorists.

"Deepfakes could alter the way we trust image and video as a society," Turek said.