Deepfake Videos Are Spreading. Can Anyone Stop Them?

Videos doctored using artificial intelligence to make them seem real, even though they are not, increasingly challenge viewers' perceptions of truth. Congress and the tech industry are taking some action to stop them, but will their efforts work?

By Michael A. Kashmer / *Digital Broadband Programming Consultant*

Think about some momentous events of the past century: Martin Luther King, Jr.'s "I have a dream" speech, the launch of Apollo 13, the destruction of the Twin Towers, to name just a few. For decades, people have relied on video and audio recordings as proof that such events really happened. But deepfakes – images and videos created using computers and machine learning software – are radically upending the notion that people can trust what they see.

Deepfakes replace people in existing images with other people's likenesses, manipulating their words and actions. Among other applications, deepfakes have been used to superimpose celebrity faces on nude bodies in pornographic videos, and to twist what political figures say and do. (Last June, a deepfake video emerged of Speaker of the House Nancy Pelosi appearing drunk and slurring her words.) Experts say deepfakes are becoming easier and more common to make and harder to detect, and, as such, pose a threat to cybersecurity and future U.S. elections.

COMPROMISED CYBERSECURITY

An October report by DeepTrace Labs reported nearly 15,000 deepfake videos online, double the number since the same time the year before. Facebook, Google and other companies that monitor deepfakes say the existing videos and images are used mostly to spread humor or fake pornography, but some experts worry that deepfakes could be used to more nefarious ends.

On the individual level, deepfake technology can be implicated in cases of sexual privacy violations, especially so-called revenge porn. Cybercriminals can use the technology to create and threaten to release compromising photos and videos of individuals if ransom demands are not met, according to cybersecurity company Forcepoint.

"At the organizational level, deepfakes will also be used to impersonate high-level targets at organizations to scam employees by transferring funds into fraudulent accounts," Alvin Rodrigues, senior director and security strategist for Asia Pacific at Forcepoint, told CNBC.

Rodrigues added, "In the political arena, we can expect deepfakes to be leveraged as a tool to discredit electoral candidates and push inaccurate falsehoods to voters via social media."

An analyst at market research company Forrester predicts that costs related to deepfake scams will exceed $250 million in 2020.

TECH, ACADEMIA SCRAMBLE FOR SOLUTIONS

As computers get more powerful and people generate ever more data, the artificial intelligence (AI) technology behind deepfakes is improving rapidly. But those same factors are simultaneously enabling AI experts to detect and take down deepfakes. It's becoming something of an arms race – will the deepfakes outpace the people trying to root them out?

Researchers in the private sector and academia are working to build systems that can automatically identify and remove deepfakes. For instance, the Boston Globe reports that DeepTrace Labs, which is developing analytical back-end systems to detect fake videos, promises a benchmark of confidence "in the high 90s." Researchers at many U.S. and international academic institutions also are working to develop AI technology to recognize altered images.

So far, they are successfully detecting current iterations of deepfakes, which have imperfections (think odd blinking or other unnatural body movements) that algorithms, if not the naked eye, can fairly easily detect. But the researchers

> Researchers are working to develop systems to automatically identify and remove deepfakes.

say improved technology will make creating fake images without such easily detectable defects possible.

"In the short term, detection will be reasonably effective," Subbarao Kambhampati, a professor of computer science at Arizona State University, told the New York Times. "In the longer term, I think it will be impossible to distinguish between the real pictures and the fake pictures."

Aleksander Madry, an associate professor of computer science at MIT focused on AI, told the Globe, "… currently this is more of a cat and mouse game where one can try to detect [deepfakes] by identifying some artifacts, but then the adversaries can improve their methods to avoid these artifacts."

Developing the right technology is just part of the solution to combat deepfakes, of course. In early January, Facebook announced it will ban deepfake videos that have been either edited or computer-generated in ways that the average person can't detect, but it will continue to allow manipulated videos that are parodies or satire. To fight the proliferation of deepfakes, more companies must revise their anything-goes policies to meet the needs of the modern era, where disinformation is becoming a norm.

LEGISLATORS TAKE ACTION

Not waiting for a technical solution, California took action to stop the spread of deepfakes by banning the distribution of "malicious" manipulated videos, audio and pictures that mimic real footage and intentionally falsify the words or actions of a political candidate, within 60 days of an election. In October, the U.S. Senate passed the Deepfake Report Act, requiring the Department of Homeland Security to assess technology used to generate deepfakes, the uses of deepfakes by foreign and domestic entities, and available countermeasures.

The bill's co-sponsor, Sen. Brian Schatz (D-Hawaii), said that "fake content can damage our national security and undermine our democracy." By directing the federal government to learn more about the scope of deepfake technology, he said, the bill is "an important first step in fighting disinformation."

PART OF LARGER TREND

Deepfakes are all the more troubling when viewed in the context of the proliferation of fake news of all types over the past few years. A 2016 Stanford study illustrated just how dire the implications may be. Middle and high school students across the country showed a "stunning and dismaying consistency" in their inability to discern factual from fake information.

The study found that more than 80 percent of middle schoolers believed that "sponsored content" was a real news story. Most high school students accepted photographs as presented without verifying them, and many high school students couldn't tell a real and a fake news source apart on Facebook. Even college students were not adept at separating fact from fiction. Most didn't suspect potential bias in a tweet from an activist group, and most Stanford students couldn't identify the difference between a mainstream and a fringe source.

Adults struggle to discern fact from fiction, too. A CNN report found that 39 percent of adults are confident that the information they are getting is accurate, and 23 percent admitted that they have shared fake news. Deepfake videos are likely to exacerbate these insidious problems. ❖

Mike Kashmer has worked in cable TV distribution, finance and programming for more than 30 years. His experience includes network startups and foreign-language programming. Reach Mike at mikekashmer@aol.com.