A deepfake of Nicolas Cage stitched onto Donald Trump. The eyes betray the deceit.

SUNY

Intelligent Machines

The Defense Department has produced the first tools for catching deepfakes

Fake video clips made with artificial intelligence can also be spotted using AI—but this may be the beginning of an arms race.

by Will Knight August 7, 2018

The first forensics tools for catching revenge porn and fake news created with AI have been developed through a program run by the US Defense Department.

Forensics experts have rushed to find ways of detecting videos synthesized and manipulated using machine learning because the technology makes it

far easier to create convincing fake videos that could be used to sow disinformation or harass people.

The most common technique for generating fake videos involves using machine learning to swap one person's face onto another's. The resulting videos, known as "deepfakes," are simple to make, and can be surprisingly realistic. Further tweaks, made by a skilled video editor, can make them seem even more real.

Video trickery involves using a machine-learning technique known as generative modeling, which lets a computer learn from real data before producing fake examples that are statistically similar. A recent twist on this involves having two neural networks, known as generative adversarial networks, work together to produce ever more convincing fakes (see "The GANfather: The man who's given machines the gift of imagination").

The tools for catching deepfakes were developed through a program—run by the US Defense Advanced Research Projects Agency (DARPA)—called Media Forensics. The program was created to automate existing forensics tools, but has recently turned its attention to AI-made forgery.

"We've discovered subtle cues in current GAN-manipulated images and videos that allow us to detect the presence of alterations," says Matthew Turek, who runs the Media Forensics program.

Tucker Carlson gets his own Nicolas Cage makeover.
UNIVERSITY AT ALBANY, SUNY

One remarkably simple technique was developed by a team led by Siwei Lyu, a professor at the State University of New York at Albany, , and one of his students. “We generated about 50 fake videos and tried a bunch of traditional forensics methods. They worked on and off, but not very well,” Lyu says.

Then, one afternoon, while studying several deepfakes, Lyu realized that the faces made using deepfakes rarely, if ever, blink. And when they do blink, the eye-movement is unnatural. This is because deepfakes are trained on still images, which tend to show a person with his or her eyes open.

Others involved in the DARPA challenge are exploring similar tricks for automatically catching deepfakes: strange head movements, odd eye color, and so on. “We are working on exploiting these types of physiological signals that, for now at least, are difficult for deepfakes to mimic,” says Hany Farid, a leading digital forensics expert at Dartmouth College.

DARPA’s Turek says the agency will run more contests “to ensure the technologies in development are able to detect the latest techniques."

The arrival of these forensics tools may simply signal the beginning of an AI-powered arms race between video forgers and digital sleuths. A key problem, says Farid, is that machine-learning systems can be trained to outmaneuver forensics tools.

Lyu says a skilled forger could get around his eye-blinking tool simply by collecting images that show a person blinking. But he adds that his team has developed an even more effective technique, but says he’s keeping it secret for the moment. “I’d rather hold off at least for a little bit,” Lyu says. “We have a little advantage over the forgers right now, and we want to keep that advantage.”

Related Video

More videos

Next-Generation Robots Need Your Help 27:36

Recommended for You

01 Once hailed as unhackable, blockchains are now getting hacked

02 Japan's Hayabusa 2 spacecraft is about to fire bullets into an asteroid

03 Russian hackers are eight times faster than North Korean groups

04 Watch a harpoon successfully spear a piece of space junk

05 Machine learning is contributing to a "reproducibility crisis" within science

More from Intelligent Machines

Artificial intelligence and robots are transforming how we work and live.

01

The Next “Deep Blue” Moment: Self-Flying Drone Racing

Spoiler: coders wanted! And a chance to win more than $2 million in cash prizes

by Lockheed Martin Vice President of Technology Strategy and Innovation, Robie I. Samanta Roy, Ph.D. (Course XVI, S.B. 1989, S.M. 1991, Ph.D. 1995)

02

The technology behind OpenAI’s fiction-writing, fake-news-spewing AI, explained

The language model can write like a human, but it doesn’t have a clue what it’s saying.

by Karen Hao

03

Self-driving cars take the wheel

Advanced technologies come together to get autonomous vehicles driving safely and efficiently.

by MIT Technology Review Insights

See details+

---

*Prices are for international subscribers.
See U.S. prices