msn

Programmers have created the ultimate tool of deception: robots that imitate people

Erica Evans

September 26, 2018

After weeks of studying “deepfakes” — computer-generated videos that depict real people saying and doing things they never did — computer science professor Siwei Lyu noticed something eerie. Finally, he figured it out: while the characters looked and sounded like real people, they didn’t blink.

So others can spot imposters online, Lyu and his team at the University of Albany developed software earlier this year that detects natural eye movements in a video. He sees his work in "digital media forensics" as essential to helping people separate reality from fiction in a world where fake news spreads faster and wider than true news.

"The potential negative impact (of fake videos) at this point, far outweighs the benefits," said Lyu. "People need to become more educated, more aware of the problem."

Deepfake technology — orginially developed for communication, educational and entertainment applications — has since been exploited to put celebrities in pornographic scenes and experts worry the convincing videos could create broader damage by undermining political campaigns, scamming people and inciting violence.

Such fraudulent videos are often made by feeding a large number of images to a computer that uses machine learning to recreate a person’s face and natural expressions. Then, the fabricated face is superimposed onto a video of someone else. Because photos with a person’s eyes closed are rare, many available programs do not replicate blinking motions well, Lyu said.

The software Lyu's lab created is one of the best available methods for identifying deepfakes, according to Lyu. In late June, Jigsaw, a spin-off initiative of Google which focusses on online security and transparency, contacted Lyu about a collaboration.

But the blinking test is already falling behind the rapidly developing technology.

To keep up, researchers are working on other fake video detection techniques that look for a human pulse or minute changes in pixel density, color or motion, said Justin Hendrix, executive director at NYC Media Lab, a non-profit which facilitates collaboration between media companies and New York City's universities.

"It's like a chess game with counter-moves," said Lyu. "We find a way to detect them, and they find a way to get around the detection."

Lyu said he has been amazed by some of the realistic videos he's seen. Most recently, in August, researchers at UC Berkeley published a new method that goes beyond cutting and pasting a person's face onto a new body and allows for full-body manipulation. A demo video shows how the team was able to project a professional dancer's moves onto another person.

But deepfakes are just one of many technologies that imitate humans and can blur the line between reality and lies, Hendrix said. With increasingly sophisticated voice imitation and text-writing bots that act like real humans, experts like Hendrix and Aviv Odavya, chief technologist at the Center for Social Media Responsibility, are trying to help the public more easily identify manipulated media, while sounding the alarm about the potential for these technologies to be used for identity theft and the proliferation of propaganda to sway public opinion.

"At this rate, we will definitely not be ready when a crisis comes," Odavya said.

The technology

Manipulating audio and video to create an alternate reality has been done for years in film.

Recent breakthroughs include computer scientists at Stanford who built a program in 2016 that combined recorded video footage with face-tracking to manipulate videos in real-time. They tested the method with images of George W. Bush, Donald Trump and Vladimir Putin. In 2017, researchers at the University of Washington, took audio clips of Barack Obama speaking and paired them with a synthesized video of Obama that matched his mouth to those words.

Equally as significant are the developments that have taken these convincing techniques from university labs and high-tech Hollywood studios and put them in the hands of the general public, said Lyu.

Now, "all users need is a computer," Lyu said.

Reddit and Github users last year started sharing their own versions of video-manipulating software on the web. In January 2018, a free desktop application called FakeApp was launched, allowing anyone to try their hand at deepfake creation. What followed was troves of silly videos featuring the likenesses of people like actor Nicholas Cage and President Donald Trump and a lot of pornography, according to news reports.

Other existing technology can replicate anyone's voice using just a few minutes of sample audio and create 3D models of faces from a few photographs.

While the technology in the wrong hands can become a tool of deception, the originators envisioned useful applications, experts said. A program that replicates someone's face and voice could be used to create footage of deceased historical figures for museums. It could be used for chat services, to fill in the gaps and make motion look natural when the Internet connection is spotty. Or, it could be used to create real-time translation — so you can Skype someone in a different county and it will appear that you are speaking each others' languages, for example.

"The technologies we're talking about are just as powerful tools for good and creating incredible news information, entertainment content and art as they are for misinformation and propaganda," said Hendrix. In his opinion, the cases of misuse are not a technology problem. "People have had the same fears around the printing press, radio, TV and film, and now social media."

Worst-case scenarios

People are used to looking at images critically because they know photos can be edited. But most of us intuitively trust a video, even if the content is surprising. We trust people's voices, too, if they are familiar to us, said Julian Sanchez, senior fellow at the Cato Institute who studies technology and privacy.

"There's a lot of companies where if someone calls IT and says 'I need you to reset my password,' they'll do that if they recognize your voice," said Sanchez. But those companies can be vulnerable to fraud with applications like Adobe Voco (which has been demo'd but not released to the public) making it easy to replicate someone's voice using just a few samples of audio, he said.

Add to that, "automated laser-phishing," a tactic that uses artificial intelligence to scan someone's email or social media posts to formulate fake, but authentic looking messages from people they know. For example, imagine getting a spam email with a harmful link that isn't from a fictitious African prince asking for money, but from a good friend you've been waiting to hear from.

Manipulated videos could be used to threaten someone with false evidence of an affair or sabotage a competitor's career with fabricated evidence of a racist comment, according to Odavya. A deep-fake video of a politician declaring war might actually start one.

Even if a video can be disproven, it might not matter.

"Very often even when stuff can be decisively refuted, the refutation doesn't travel as quickly as the initial scandalous claim," said Sanchez. "In a political campaign context, if something scandalous drops in late October, it might not matter that experts can show it's a fake two weeks later."

The ramifications may ripple beyond the parties involved and irreparably damage the reputation of a publication that inadvertently releases something fake. "We've seen now for a long period of time, diminishing trust in media. People worry now that deepfakes will be the final nail in the coffin," said Hendrix.

While teams like Lyu's in Albany are coming up with tools to more easily identify deepfakes and other deception practices, common users need to start critically thinking about the media they consume, Sanchez added. "Seeing is not believing."

"Instead of just pushing retweet or the share button, do a 30-second check of where the video came from and look at the contents," said Lyu. "If we are all diligent and understand that media can be manipulated, we can contribute to defer this wave of fake media."

Robot vs. human

Earlier this year, Google unveiled a new product called Duplex which has the ability to call and make an appointment for you. It says “um” and “mmhmm” and phrases like “gotcha, thanks.” It hesitates and pauses like a human would, and as a result, when Google tested it, people on the other end of the line couldn’t tell they were talking to a robot.

But the criticism was immediate, prompting Google to assure users would be notified that they are talking to the disembodied voice of Google’s automated booking service.

“We believe that providing transparency and control are very important when developing new technologies, and from the early days of building the system we’ve thought carefully about how we can incorporate this,” said a press release provided by Google spokesperson Joshua Cruz. Google is moving forward, testing the product with a select group of businesses, the statement said.

But these assurances haven't eased everyone’s fears.

California legislators have passed a law that requires "bots" in certain settings to disclose that they are not people. And U.S. Sen. Dianne Feinstein, D-Calif., has introduced a similar bill in the U.S. Senate.

Sen. Mark Warner, D-Va., vice chairman of the U.S. Senate Select Committee on Intelligence, recently published an interim white paper that argued public understanding of technology that imitates humans is crucial and more needs to be invested in tech and media literacy.

“There are no easy solutions,” said Sanchez, of the Cato Institute.

In addition to promoting public awareness, Sanchez recommends software engineers incorporate watermarks to easily identify altered media. Symbols could appear on a video that’s been fabricated, or a chime could sound periodically over synthesized audio.

Odavya said social media platforms should take the lead on identifying bots and fake videos.

“Facebook and Twitter have a responsibility ensure that their sites do not reward misinformation and sensationalism over informative and accurate content,” said Odavya, noting not much has been done on that front.

Although Google's Jigsaw reached out to Lyu about a collaboration in June, he hasn’t been contacted by any social media companies.

“There has been minimal public engagement by platforms on this issue," Odavya said. "Some of the platforms may be taking baby steps in the right direction, but this is a small fraction of the level of research, design, and development that are needed immediately if we want to get ahead of this threat.”