Deepfakes: More Frightening Than Photoshop on Steroids

By Judge Herbert B. Dixon Jr. (Ret.)

ASSOCIATED PRESS PHOTO

Seeing is believing, so they say. However, any truism associated with this ancient statement may disappear from our ethos. Have you heard about deepfakes?

A deepfake is video created or altered using digital means with the aid of artificial intelligence (AI). With deepfakes, persons appear to do or say things that did not happen. It is sometimes difficult for an expert to distinguish a deepfake from an unaltered video. In other words, the world has turned on its head with a new governing principle—"Do not necessarily believe what you see!" The dangers presented by deepfakes are more frightening than Photoshop on steroids. How did we get here?

The term *deepfakes* comes from the name of a Reddit contributor who surprised the technology community in 2017 when, using publicly available AI-driven software, he successfully stitched or imposed the faces of celebrities onto the bodies of people in pornographic videos. The resulting videos were shockingly realistic. Although a few discriminating viewers could spot imperfections or telltale signs of fraud, the resulting videos were nevertheless unsettling—especially to the victims whose identities were falsely depicted in the pornographic productions. The AI-driven software used to create the phony videos employed a concept known as deep learning to accomplish the deception. Let me give a few simplified descriptions of the technology used to create these deceptive videos.

Merely Slowing or Speeding a Video Does Not Create a Deepfake

It takes more than a rudimentary modification of a video to create a deepfake. Some would use the term "shallow fake" or "cheap fake" to describe a minor adjustment such as slowing down or speeding up a video, but this type of video alteration can still deceive. Do you recall the spring 2019 video circulating of House Speaker Nancy Pelosi slurring her words? Many people who saw the video spread misinformation on their social media accounts of their belief that Pelosi was drunk, was the victim of a stroke, or had suffered some other loss of mental acuity.[1]

A 2018 incident that went viral on social media involves a video of CNN White House Reporter Jim Acosta moving his body and arms to avoid a White House intern's effort to physically take away a microphone from Acosta's possession. Some enterprising video editor, apparently not a fan of Acosta, deleted Acosta's statement of "Pardon me, ma'am" and sped up the video during that portion of the incident, making it appear that Acosta's actions were physically aggressive against the intern.[2]

Judge Herbert B. Dixon Jr. retired from the Superior Court of the District of Columbia after 30 years of service. He is a former chair of both the National Conference of State Trial Judges and the ABA Standing Committee on the American Judicial System and a former member of the Techshow Planning Board. You can reach him at Jhbdixon@gmail.com. Follow Judge Dixon on Twitter @Jhbdixon.

Bringing a Single Image to Life Shows the Capability of Deepfake Technology

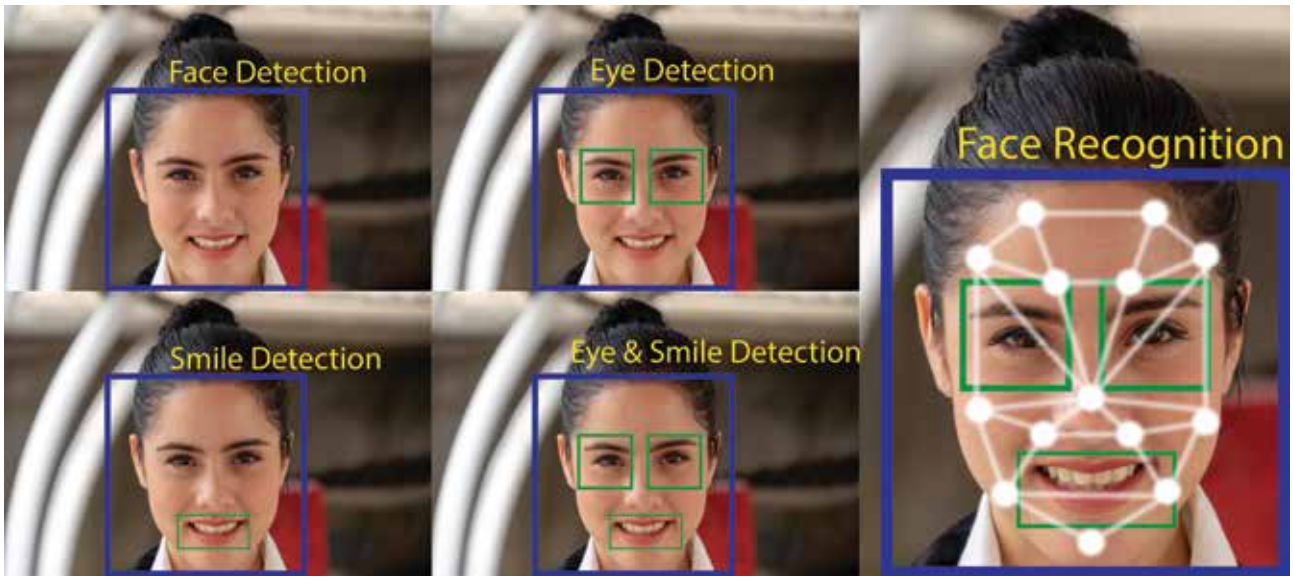
Researchers in Russia (where else but Russia?) at Moscow's Samsung AI Center and Skolkovo Institute of Science and Technology used AI-driven software to animate single images, including the *Mona Lisa* painting by the Italian Renaissance artist Leonardo da Vinci, the 1665 oil painting of *Girl with a Pearl Earring* by Dutch Golden Age artist Johannes Vermeer, and the 1883 *Portrait of an Unknown Woman* by Russian artist Ivan Kramskoi. These efforts went far beyond the work of cartoon animators. The researchers produced a paper entitled "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models"[3] describing the technology. The researchers also created a video[4] that shows the results of their work making the iconic still images move and talk as if they were real people.

Creating an Ultra-Realistic Deepfake

The phrase "ultra-realistic deepfake" is a misnomer because, even absent a generally accepted definition, a deepfake is a deceptive video showing a subject doing or saying things that did not occur. Although Hollywood for years has used highly skilled artists to alter video footage, deepfake technology is a new paradigm—a game changer. The product of today's deepfake technology has become unnervingly realistic. Because of the advances in deepfake technology, members of the U.S. Congress have requested a formal report from the Director of National Intelligence.[5] In addition, the military, through the Defense Advanced Research Projects Agency (DARPA), has embarked on research efforts with numerous academic and corporate institutions to develop technology that can detect deepfakes.[6] A concern often expressed by the researchers is that continuing advances in deepfake technology may render it impossible for one AI-driven system to detect a video created or modified by another AI-driven system.

Creating a deepfake involves modifying images or videos using a machine-learning technique called a "generative adversarial network" (GAN). The AI-driven software detects the way a subject moves his or her mouth and face from the source images and duplicates those movements on the subject of another video. To understand how a deepfake product is improved, imagine that a deepfake creator has a substantial database of videos and two AI-driven software packages. One software package trains itself on the database of source images and creates video forgeries. The second AI-driven software package detects forgeries. Through a reiterative process, the first software package creates and modifies fake videos until the other software package can no longer detect the forgery. This is known as "unsupervised learning"—when machine-language models teach themselves. The larger the database of source images, the easier it is for the deepfake creator to produce a convincing deepfake. For this reason, videos of high-profile politicians and Hollywood celebrities have been used frequently to create deepfakes, as there is a substantial supply of publicly available video footage to train the deepfake creating system.

Said another way, the deepfake-creating software studies the statistical patterns in a data set, such as a set of images or videos, and then generates convincing fake videos.[7] The deepfake detection software then tries to distinguish between real and fake examples. Multiple iterations of the process enable the deepfake-creating software to produce a result that is increasingly difficult for the deepfake detection software to detect. Because the creating AI-driven system is designed to produce a deepfake that the detecting AI-driven system is unable to distinguish, some commentators suggest the eventuality that the technology to detect deepfakes may always be in catch-up mode.[8]



Computer forensics experts say that judges and litigators are not thinking enough about how to address deepfake evidentiary issues.

AI-Driven Software Can Create Images of Fake People, Change Facial Expressions, and Imitate Voices

Although the term *deepfake video* has been used throughout this column, audio forgeries are often a part of the process. An AI startup in Montreal, Canada, has demonstrated the ability to synthesize a person's voice with just a one-minute recording of the original. Adobe has come up with the unique name of "Photoshop for audio." Its software technology, "VoCo," requires merely a 20-minute sample of someone's voice, which the AI-driven software analyzes and learns to mimic. After that, the Photoshop AI-driven software will produce the words in the targeted person's voice from a typed statement.[9] Even more, AI-driven software using machine-learning algorithms has created fake images of faces that don't belong to real people and modified images by changing a frown into a smile, night into day, and summer into winter.

Using Deepfakes to Create Mischief

Political pundits say that deepfakes may cause chaos during the 2020 presidential elections. Business consultants say deepfakes will undermine future business ventures. Computer forensics experts say that judges and litigators are not thinking enough about how to address deepfake evidentiary issues when they show up in a few years. Consider the following possibilities.

What would happen if, just before an election, a deepfake video surfaces of a political candidate supporting a policy opposed to the candidate's primary platform or speaking of the candidate's deceptive embrace of a policy he does not support. The video could cause serious harm to the candidate's electability.

Consider the possible adverse influence on a shareholder's meeting called to consider a major business acquisition if, just before the meeting, a deepfake video surfaces of the CEO musing about his upcoming riches from the merger and the sucker shareholders who do not understand what is going on.

Lastly, judges and lawyers, consider what actions should be taken during a pretrial conference where (1) a party offers an exhibit of a cell phone video disclosed during discovery that supports the offering party's position of an agreement reached by the two parties, (2) the offering party will testify affirmatively concerning the authenticity and accuracy of the video, and (3) the opposing party will testify that he never said the words portrayed in the video.

With the constant advancements in deepfake technology, all of the above examples are plausible. Numerous experts on this subject agree that deepfakes created by AI-driven software can be nearly impossible to detect and the number of fake videos will grow with advances in free, readily available tools to create them.

Final Thoughts

In today's climate of "fake news" claims and increasing awareness of forged signatures and "photoshopped" images, it is possible that the public may be naturally skeptical of certain types of evidence and fully prepared to consider claims that a video does not accurately reflect what the subject of the video actually said and did. What is the answer to this dilemma if technology provides no definitive method to detect deepfakes? Will our society evolve to a new paradigm in which no video is accepted as proof of what it purports to show? Will the phrase "seeing is believing" cease to exist? "The proof is in the pudding"—whatever that means. ■

Endnotes

1. Sarah Mervosh, *Distorted Videos of Nancy Pelosi Spread on Facebook and Twitter, Helped by Trump*, N.Y. Times (May 24, 2019), https://nyti.ms/2ZvJ8V5.

2. Drew Harwell, *White House Shares Doctored Video to Support Punishment of Journalist Jim Acosta*, Wash. Post (Nov. 8, 2018), https://wapo.st/2x6iaXW.

3. Egor Zakharov et al., *Few-Shot Adversarial Leaning of Realistic Neural Talking Head Models*, https://bit.ly/30NtiGW.

4. Egor Zakharov et al., *Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*, YouTube (May 21, 2019), https://bit.ly/2WnVRer.

5. Donie O'Sullivan, *When Seeing Is No Longer Believing*, CNN Bus. (2019), https://cnn.it/2DCzQhJ.

6. Will Knight, *The US Military Is Funding an Effort to Catch Deepfakes and Other AI Trickery*, MIT Tech. Rev. (May 23, 2018), https://bit.ly/2IYzLIL.

7. *Id.*

8. See, for example, statements by David Gunning, DARPA project manager, reported in *id.*, and in J.M. Porup, *How and Why Deepfake Videos Work—and What Is at Risk*, CSO (Apr. 10, 2019), https://bit.ly/2Pmk9jc.

9. Harmon Leon, *Why AI Deepfakes Should Scare the Living Bejeezus Out of You*, Observer (June 12, 2019), https://bit.ly/2Zp4kMu; Scott Amyx, *Using Artificial Intelligence for Forgery: Fake Could Be Eerily Real*, IoT Agenda (July 26, 2017), https://bit.ly/2X146cV.