HUFFPOST

POLITICS

06/12/2019 04:03 pm ET **Updated** Jun 12, 2019

Deepfake Videos And The Threat Of Not Knowing What's Real

The rapid spread of technology that can essentially bend our digital reality has alarmed experts worldwide.

By Jesselyn Cook

It's November 2020, only days before the presidential election. Early voting is underway in several states as a video suddenly spreads across social media. One of the candidates has disclosed a dire cancer diagnosis, and is making an urgent plea: "I'm too sick to lead. Please, don't vote for me." The video is quickly revealed to be a computer-generated hoax, but the damage is done — especially as trolls eagerly push the line that the video is actually real, and the candidate has just changed her mind.

Such a scenario, while seemingly absurd, would actually be possible to achieve using a "deepfake," a doctored video in which a person can be made to appear as if they're doing and saying anything. Experts are issuing increasingly urgent warnings about the advance of deepfake technology — both the realistic nature of these videos, and the ease with which even amateurs can create them. The possibilities could bend reality in terrifying ways. Public figures could be shown committing scandalous acts. Random women could be inserted into porn videos. Newscasters could announce the start of a nonexistent nuclear war. Deepfake technology threatens to provoke a genuine civic crisis, as people lose faith that anything they see is real.

House lawmakers will convene on Thursday for the first time to discuss the weaponization of deepfakes, and world leaders have begun to take notice.

“People can duplicate me speaking and saying anything. And it sounds like me and it looks like I’m saying it — and it’s a complete fabrication,” former President Barack Obama said at a recent forum. “The marketplace of ideas that is the basis of our democratic practice has difficulty working if we don’t have some common baseline of what’s true and what’s not.” He was featured in a viral video about deepfakes that portrays him calling his successor a “total and complete dipshit.”

How Deepfakes Are Made

Directors have long used video and audio manipulation to trick viewers watching scenes with people who didn’t actually participate in filming. Peter Cushing, the English actor who played “Star Wars” villain Grand Moff Tarkin before his death in 1994, reappeared posthumously in the 2016 epic “Rogue One: A Star Wars Story.” “The Fast and the Furious” star Paul Walker, who died before the series’ seventh movie was complete, still appeared throughout the film through deepfake-style spoofing. And showrunners for The Sopranos had to create scenes with Nancy Marchand to close her storyline as Tony’s scornful mother, after Marchand died between the second and third seasons of the show.

Thanks to major strides in the artificial intelligence software behind deepfakes, this kind of technology is more accessible than ever.

Here’s how it works: Machine-learning algorithms are trained to use a dataset of videos and images of a specific individual to generate a virtual model of their face that can be manipulated and superimposed. One person’s face can be swapped onto another person’s head, like this video of Steve Buscemi with Jennifer Lawrence’s body, or a person’s face can be toyed with on their own head, like this video of President Donald Trump disputing the veracity of climate change, or this one of Facebook CEO Mark Zuckerberg saying he “controls the future.” People’s voices can also be imitated with advanced technology. Using just a few minutes of audio, firms such as Cambridge-based Modulate.ai can create “voice skins” for individuals that can then be manipulated to say anything.

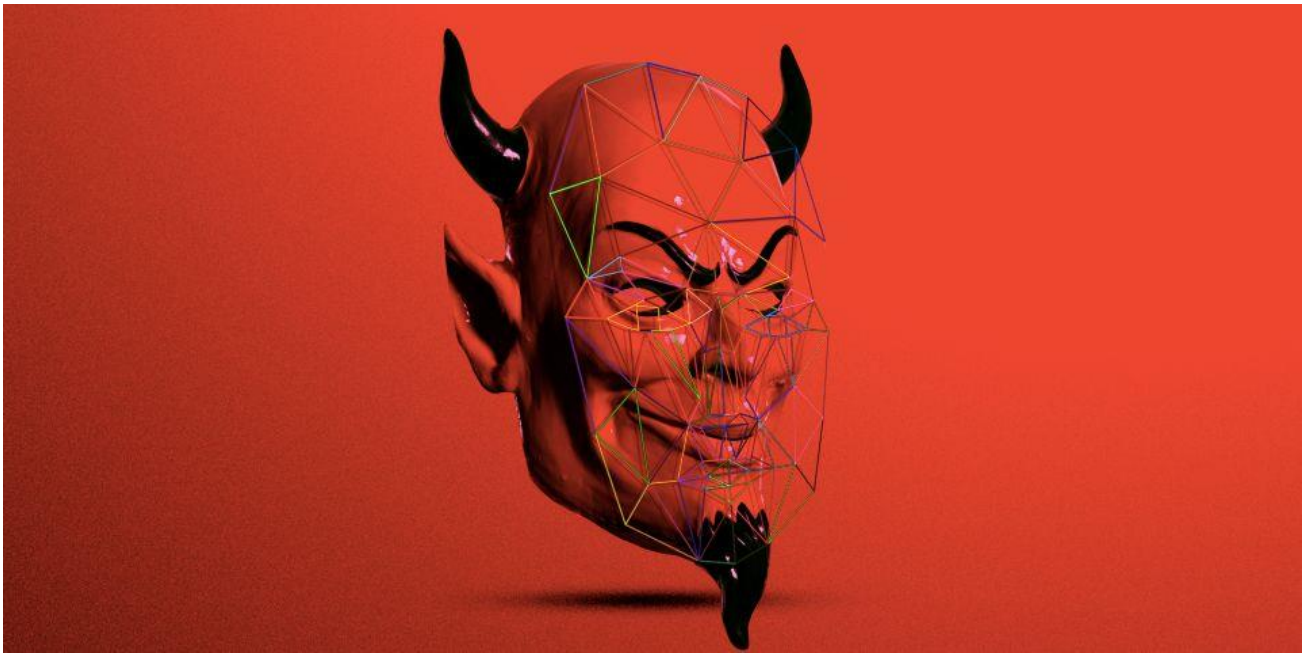
It may sound complicated, but it's rapidly getting easier. Researchers at Samsung's AI Center in Moscow have already found a way to generate believable deepfakes with a relatively small dataset of subject imagery — "potentially even a single image," according to their recent report. Even the "Mona Lisa" can be manipulated to look like she's come to life:

There are also free apps online that allow ordinary people with limited video-editing experience to create simple deepfakes. As such tools continue to improve, amateur deepfakes are becoming more and more convincing, noted Britt Paris, a media manipulation researcher at Data & Society Research Institute.

"Before the advent of these free software applications that allow anyone with a little bit of machine-learning experience to do it, it was pretty much exclusively entertainment industry professionals and computer scientists who could do it," she said. "Now, as these applications are free and available to the public, they've taken on a life of their own."

The ease and speed with which deepfakes can now be created is alarming, said Edward Delp, the director of the Video and Imaging Processing Laboratory at Purdue University. He's one of several media forensics researchers who are working to develop algorithms capable of detecting deepfakes as part of a government-led effort to defend against a new wave of disinformation.

"It's scary," Delp said. "It's going to be an arms race."



NICOLAS ORTEGA FOR HUFFPOST

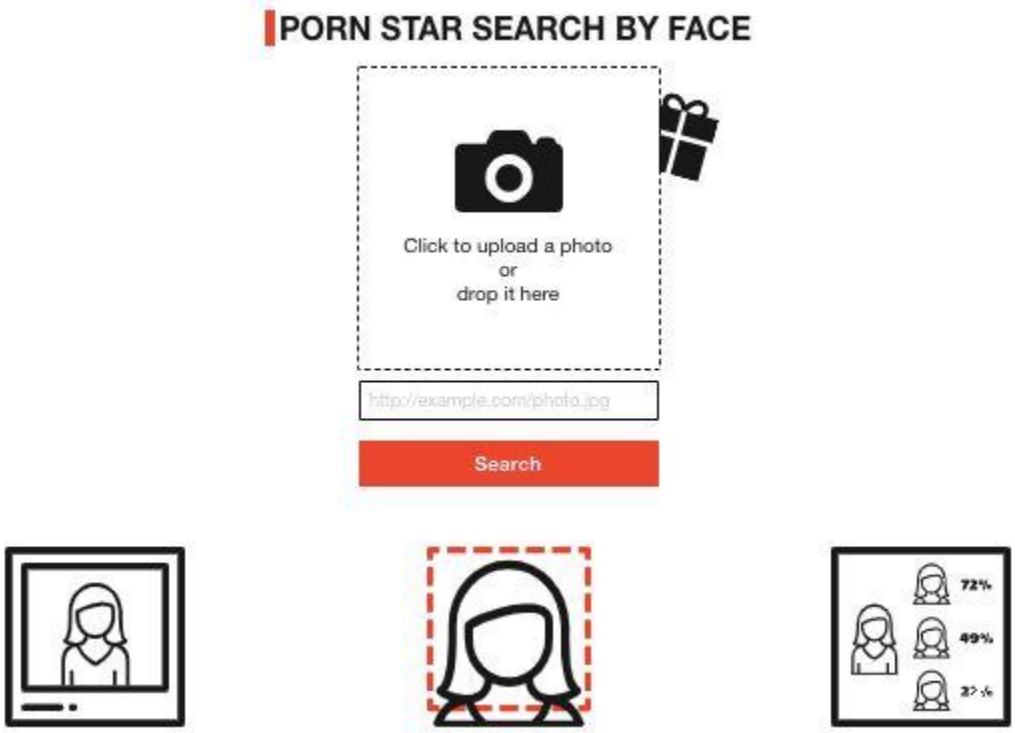
Much of the discussion about the havoc deepfakes could wreak remains hypothetical at this stage — except when it comes to porn.

Videos labeled as “deepfakes” started in porn. The term was coined in 2017 by a Reddit user who posted fake pornographic videos, including one in which actor Gal Gadot was portrayed to be having sex with a relative. Gadot’s face was digitally superimposed onto a porn actor’s body, and apart from a bit of glitching, the video was virtually seamless.

“Trying to protect yourself from the internet and its depravity is basically a lost cause,” actor Scarlett Johansson, who’s also been featured in deepfake porn videos, including some with millions of views, told The Washington Post last year. “Nothing can stop someone from cutting and pasting my image.”

It’s not just celebrities being targeted — any person with public photos or videos clearly showing their face can now be inserted into crude videos with relative ease. As a result, revenge porn, or nonconsensual porn, is also becoming a broadening threat. Spurned creeps don’t need sex tapes or nudes to post online anymore. They just need pictures or videos of their ex’s face and a well-lit porn video. There are even photo search engines (which

HuffPost won't name) that allow a person to upload an image of an individual and find a porn star with similar features for optimal deepfake results.



Screenshot from a reverse image search engine.

In online deepfake forums, men regularly make anonymous requests for porn that's been doctored to feature women they know personally. The Post tracked down one woman whose requestor had uploaded nearly 500 photos of her face to one such forum and said he was "willing to pay for good work." There's often no legal recourse for those who are victimized by deepfake porn.

Beyond the concerns about privacy and sexual humiliation, experts predict that deepfakes could pose major threats to democracy and national security, too.

American adversaries and competitors "probably will attempt to use deep fakes or similar machine-learning technologies to create convincing — but false — image, audio, and video files to augment influence campaigns directed against the United States and our allies and partners," according to

the 2019 Worldwide Threat Assessment, an annual report from the director of national intelligence.

Deepfakes could be deployed to erode trust in public officials and institutions, exacerbate social tensions and manipulate elections, legal experts Bobby Chesney and Danielle Citron warned in a lengthy report last year. They suggested videos could falsely show soldiers slaughtering innocent civilians; white police officers shooting unarmed black people; Muslims celebrating ISIS; and politicians accepting bribes, making racist remarks, having extramarital affairs, meeting with spies or doing other scandalous things on the eve of an election.

“If you can synthesize speech and video of a politician, your mother, your child, a military commander, I don’t think it takes a stretch of the imagination to see how that could be dangerous for purposes of fraud, national security, democratic elections or sowing civil unrest,” said digital forensics expert Hany Farid, a senior adviser at the Counter Extremism Project.

The emergence of deepfakes brings not only the possibility of hoax videos spreading harmful misinformation, Farid added, but also of real videos being dismissed as fake. It’s a concept Chesney and Citron described as a “liar’s dividend.” Deepfakes “make it easier for liars to avoid accountability for things that are in fact true,” they explained. If a certain alleged pee tape were to be released, for instance, what would stop the president from crying “deepfake”?

Alarm Inside The Federal Government

Though Thursday’s congressional hearing will be the first to focus specifically on deepfakes, the technology has been on the government’s radar for a while.

The Defense Advanced Research Projects Agency, or DARPA, an agency of the U.S. Department of Defense, has spent tens of millions of dollars in recent years to develop technology that can identify manipulated videos and images, including deepfakes.

Media forensics researchers across the U.S. and Europe, including Delp from Purdue University, have received funding from DARPA to develop machine-

learning algorithms that analyze videos frame by frame to detect subtle distortions and inconsistencies, to determine if the videos have been tampered with.

We might get to a situation in the future where you won't want to believe an image or a video unless there's some authentication mechanism. Edward Delp, director of the Video and Imaging Processing Laboratory at Purdue University

Much of the challenge lies in keeping pace with deepfake software as it adapts to new forensic methods. At one point, deepfakes couldn't incorporate eye-blinking or microblushing (facial blushing that's undetectable to the naked eye), making it easy for algorithms to identify them as fake, but that's no longer the case.

"Our method learns all these new attack approaches so we can then detect those," Delp said.

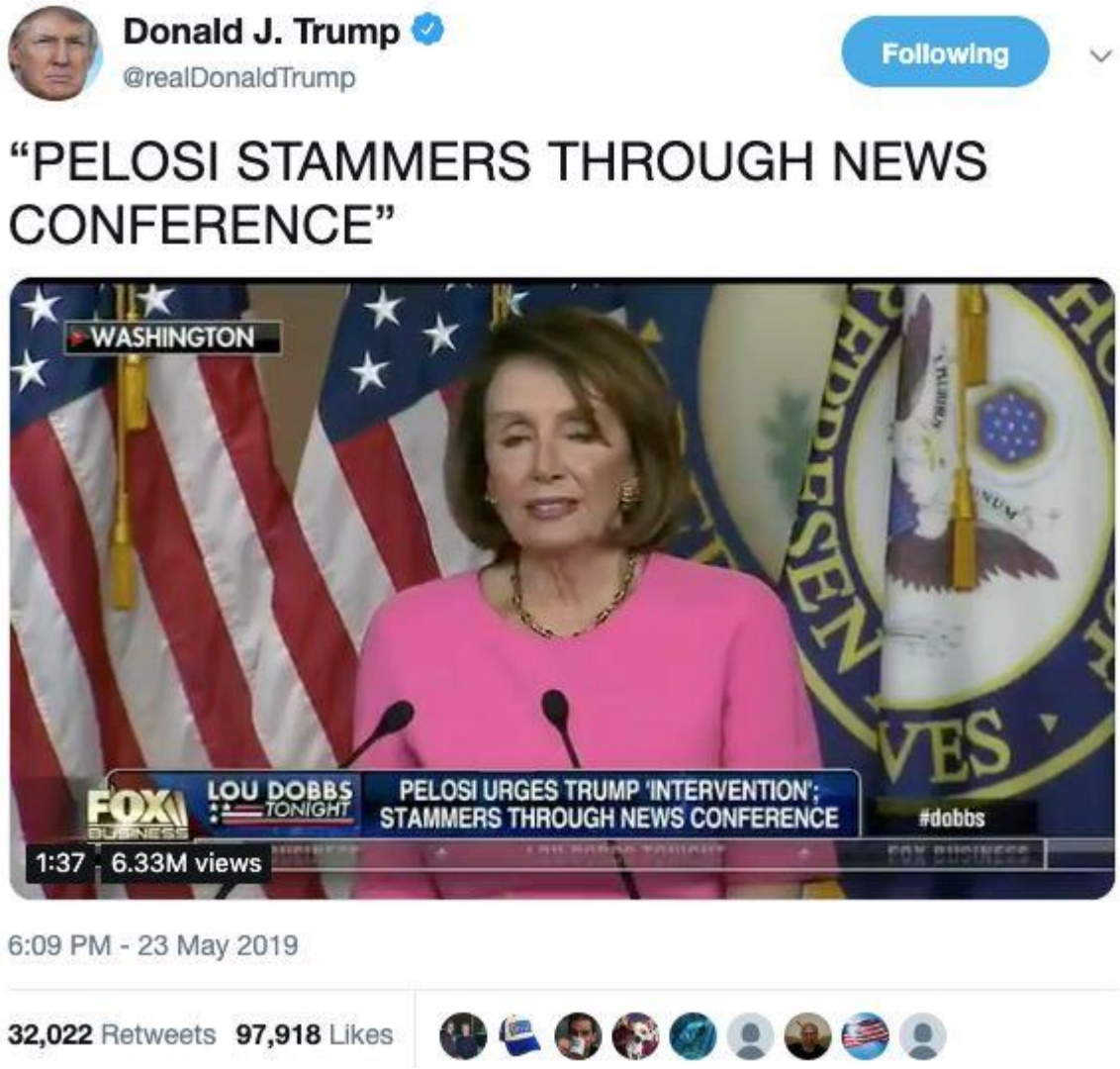
"As the people making these videos get more and more sophisticated with their tools, we're going to have to get more and more sophisticated with ours," he added. "We might get to a situation in the future where you won't want to believe an image or a video unless there's some authentication mechanism."

With a presidential election on the horizon, politicians have also started to sound the alarm about deepfakes. Congress introduced the Malicious Deep Fake Prohibition Act in December, which would make it illegal to distribute deepfakes with an intent to "facilitate criminal or tortious conduct," and the Algorithmic Accountability Act in April, which would require tech companies to audit their algorithms for bias, accuracy and fairness.

"Now we have deepfake technology, and the potential for disruption is exponentially greater," Rep. Adam Schiff (D-Calif.) said last month at a panel event in Los Angeles. "Now, in the weeks leading up to an election, you could have a foreign power or domestic party introduce into the social media bloodstream a completely fraudulent audio or video almost indistinguishable from real."

A New Breed Of ‘Fake News’

Despite Trump’s countless tirades against fake news, his own administration has shared hoax videos online. The president himself has circulated footage that was manipulated to deceive the public and stoke partisan tensions.



In May, Trump tweeted a montage of clips featuring Nancy Pelosi, the Democratic speaker of the House of Representatives, that was selectively edited to highlight her verbal stumbles.

“PELOSI STAMMERS THROUGH NEWS CONFERENCE,” Trump wrote in his tweet, which he has yet to delete. His attorney Rudy Giuliani also tweeted a link to a similar video, with the text: “What is wrong with Nancy Pelosi? Her

speech pattern is bizarre." That video, as it turns out, had been carefully tampered with to slow Pelosi's speech, giving the impression that she was intoxicated or ill.

Months earlier, White House press secretary Sarah Huckabee Sanders tweeted a video that had been altered in an attempt to dramatize an interaction between CNN reporter Jim Acosta and a female White House intern.

The video, which Sanders reportedly reposted from notorious conspiracy theorist Paul Joseph Watson, was strategically sped up at certain points to make it look as if Acosta had aggressively touched the intern's arm while she tried to take a microphone away from him.

"We will not tolerate the inappropriate behavior clearly documented in this video," Sanders wrote in her tweet, which she, too, has yet to delete.

Neither the video of Pelosi nor the one of Acosta and the intern was a deepfake, but both demonstrated the power of manipulated videos to go viral and sway public opinion, said Paris, from Data & Society Research Institute.

"We're in an era of misinformation and fake news," she said. "People will believe what they want to believe."

When Hoaxes Go Viral

In recent years, tech giants have struggled — and sometimes refused — to curb the spread of fake news on their platforms.

The doctored Pelosi video is a good example. Soon after it was shared online, it went viral across multiple platforms, garnering millions of views and stirring rumors about Pelosi's fitness as a political leader. In the immediate aftermath, Google-owned YouTube said it would remove the video, but days later, copies were still circulating on the site, CNN reported. Twitter declined to remove or even comment on the video.

Facebook also declined to remove the video, even after its third-party fact-checkers determined that the video had indeed been doctored, then doubled down on that decision.

“We think it’s important for people to make their own informed choice about what to believe,” Facebook executive Monika Bickert told CNN’s Anderson Cooper. When Cooper asked if Facebook would take down a video that was edited to slur Trump’s words, Bickert repeatedly declined to give a straight answer.

“We aren’t in the news business,” she said. “We’re in the social media business.”

More and more people are turning to social media as their main source for news, however, and Facebook profits off the sharing of news — both real and fake — on its site.

Even if the record is corrected, you can’t put the genie back in the bottle. Digital forensics expert Hany Farid

Efforts to contain deepfakes in particular have also had varying levels of success. Last year, Pornhub joined other sites including Reddit and Twitter in explicitly banning deepfake porn, but has so far failed miserably to enforce that policy.

Tech companies “have been dragging their feet for way too long,” said Farid, who believes the platforms should be held accountable for their role in amplifying disinformation.

In an ecosystem where hoaxes are so often designed to go viral, and many people seem inclined to believe whatever information best aligns with their own views, deepfakes are poised to bring the threat of fake news to a new level, added Farid. He fears that the tools being designed to debunk deepfakes won’t be enough to reverse the damage that’s caused when such videos are shared all over the web.

“Even if the record is corrected, you can’t put the genie back in the bottle.”

How To Spot A Deepfake

The experts who spoke to HuffPost provided some tips to help you spot amateur deepfakes yourself:

Watch the perimeter of the face: Inconsistencies including discoloration and blurring in this area could mean one person's face has been swapped onto another person's head.

Analyze eye-blinking: Deepfake visualizations of people sometimes blink less often than a real person would, or don't blink at all.

Close your eyes and listen carefully: If the video shows a public figure, focus on their voice. Does it actually sound like that person speaking?

Watch on your computer, not your phone: Tell-tale signs of deepfakes, such as glitching or pixelation, may not be clearly visible on smaller screens.

Use common sense and ask basic questions: Does what you're watching make sense? Who first shared the video, and why? As we approach a time when you literally may not be able to believe your eyes, vigilance is crucial.

HUFFPOST

https://www.huffpost.com/entry/deepfake-videos-and-the-threat-of-not-knowing-whats-real_n_5cf97068e4b0b08cf7eb2278?fbclid=IwAR1FwrotW0aHV5KddWrEB4Kzy7prhUYOMNuWu-7GljICv_WzXWJX2m3xxFM&guccounter=1&guce_referrer=aHR0cHM6Ly9sLmZhY2Vib29rLmNvbS8&guce_referrer_sig=AQAAAN4ZdhureUdgciiEG4RCMn9dewUPv8n2t3wRwkP86PAmGtxJVWr3BskwR0zlpO6sj8qba7-shlgX-eKRJfjUWo0z025Xe5GL5Ln5Jr9OpYKf5-3mXn_3zm8YHTFP6ghLhycepkypIkxDqvIkYB7UwOQMyEhkXmFn_F0G8pO7PElc