

Detecting Deepfakes : Strategies and Tools

by Donald Harvey Marks

Physician scientist and 3rd generation veteran

Deepfake videos are a new concern as they can manipulate audiences with a high success rate. It can be hard to detect them and the technology is not ready enough yet, but, we have to try anyway.

Videos, usually, are considered to be the ultimate proof of something. They are often regarded as the final tool to establish a claim, because, they are thought, at least until a few years past, to be undeniable, unalterable, and authentic. But, as the technology advances with its own pace, videos can be shrouded with lies now. The technology of deepfake made it possible already.

What is a deepfake?

[Deepfakes](#), also known as deepfake videos, are the kind of videos that are created from still images, sample videos with the help of artificial intelligence (AI). A deepfake video is only considered to be a threat when it is used to alter reality, spread lies, create chaos, or make false statements by a character.

Deepfake technology is a growing and latest branch of artificial intelligence. Deepfake does not always mean to be a rogue technology. With the [help of deepfake](#), amazing things can be done. You can even have Albert Einstein teach you his revolutionary theory of relativity with his own voice in a video lecture. Well, how can we expect to have only good things from such a smart technology! Somewhere, someone with an evil-mind, ought to be thinking otherwise.

Detecting deepfakes : Instinct v AI

Artificial intelligence is regarded as a blessing for modern civilization, but, for some, it has always been a competitor to human beings' intelligence, biological instincts. There are questions and arguments if AI is going to outsmart us to any extent in future. Science-fiction films like *Terminator* or *Eagle Eye* showed great enthusiasm to depict how dangerous and smart AI can be. But, are they really going to be that clever to fool us and take over the world?

We are too early to reach a verdict on that. The first thing about detecting a deepfake content is what your instinct says. It's kind of an AI versus instinct game and some experts still keep faith in human instinct.

When [Shelly Palmer](#), a technology consultant, was asked what is the number one thing people should look out for on social media to detect a deepfake, he replied, "Oddly enough, it's more than common sense." Although, he admitted that they are getting harder and harder to detect, he suggested to look critically at what you're looking at. Your instinct is your primary defence.

It's all about controlling your confirmation biases. Do you want to just believe or think twice? Listen to your instincts, does this content look like it's weird to be existing? Or, you just don't care and accept what is being presented in front of you?

How to detect deepfake pictures

Deepfake images are created using the generative adversarial network (GAN) technology. In the GAN process, two programs work together to create fake images. One of these programs creates the doctored photo from a lot of several sample photos of real people. The other program tries to sniff out the fakeness. If the second program fails, then a deepfake is born.

[Photos from GAN technology](#) look flawless at first glance but, they sure have their glitches. A deepfake image can be detected by closely looking at it, some glitches will be visible. The shade, the borders, overlapping materials near the teeth, eyes, ear etc are some possible factors to detect a GAN image.

The background of the image may look fuzzy too. Even the dress might appear in a sort of different colour in different parts. Looking at them closely and taking time with the observation can end up with success.

How to detect deepfakes using machine learning

Computer systems are created with algorithms, programs full of codes to execute specific commands. Algorithms are primarily generated by humans but, in some advanced cases, algorithms can learn and adapt to new data by them, meaning no human interaction. This method is known as machine learning.

The machine learns by itself; that's pretty impressive, isn't it? So, how about we ask the machine to learn about deepfakes, evaluate and analyze the contents, and then tell us how to detect them? That sounds quite plausible.

[Facebook](#) figured this out and already started working on it. With the partnership of Microsoft, Amazon, [Partnership on AI](#), and academics from universities, it has created, of course ethically, [a lot of deepfake videos](#). To be specific, the number of the amount crosses 100k. This program is called the [Deepfake Detection Challenge Dataset](#). These deepfake videos will be fed to its deepfake detection AI program, and the AI will use its machine learning method to improve the detection-process.

Facebook is simply using deepfakes to detect deepfakes. The social media platform has also arranged a competition, [Deepfake](#)

[Detection Challenge](#), to test its dataset. 2,114 participants submitted their 35,000 models and Selim Seferbekov won the top position. His model was 65% accurate to detect a video whether it was a deepfake or not.

Will this do any good in near future, the question remains. [Hao Li, a computer scientist](#) and an associate professor at the University of Southern California reckons deepfake detection algorithms are not going to stand the fight for a long enough time. He opines at some point no deepfake detector is going to be able to detect anything.

Hao Li and his team [designed a deepfake detector](#) that works based on an individual's bodily movement, facial gestures etc. But, Li seems to think that this invention is also going to be obsolete in future, since the technology is evolving with a virus / anti-virus dynamic.

Microsoft Video Authenticator

Everybody would adore an easier way, a tool, that would help detect deepfake. Microsoft came up with the [Video Authenticator tool](#) in September 2020, hoping to detect fake contents, both photos and videos, by analyzing technical signs that are most likely not visible to human eyes. As it describes in its blog post, the tool works “by detecting the blending boundary of the deepfake and subtle fading or greyscale elements that might not be detectable by the human eye.”

To build the tool, the company used machine learning techniques. Experts think this disinformation-combatant tool may become outdated too, just like every other detection system out there. But, good news is, Microsoft is working on another system that will help content producers add [hidden code](#) to their videos. This code will help pointing out any kind of changes if made later on.

Reality Defender

[Reality Defender](#) is developed by Supasorn Suwajanakorn, a Google engineer and a computer scientist. In a [Ted Talk](#) in 2018, he announced the tool, which works as a browser [plugin and scans for](#) fake contents.

Sensity

[Sensity](#) is a free deepfake detection tool we found out on the Internet while researching deepfakes. The tool analyze suspicious files and URLs to detect types of AI-generated visual threats. We tested the tool by using a GAN-generated image, which we obtained from the website [thispersondoesnotexist](#). For context, this site creates images of people who actually don't exist. The process is done by style-based GAN architecture, commonly known as [StyleGAN](#).

The result was satisfactory, as it labeled the face as a GAN-generated face with a confidence scale of 99.9%. There was a result-box of face swap, and the result was negative for that, because the photo was not swapped with any other face.

The tool accepts only images and videos. The accepted formats of images are: png, jpeg, jfif, tiff; mp4, mov. For videos, it has some limits like 30 MB, 10 minutes, 1440p.

Deepware

[Deepware](#) scans and detects deepfake videos. In its scanner-box, all you need to do is paste the video link and it will show the result after processing it. If you don't have any URL (Uniform Resource Locator), simply upload the video.

We tested the tool with a [deepfake video](#) first, and then an [authentic one](#). The result made sense.

Wrapping up

Combatting deepfakes is really necessary, and it's not always about just technology and tools to fight the fake. [Government regulations and legislative procedures](#) are needed to fight this terror. People behind deepfakes with bad intention ought to be brought under punishments by the governments and law enforcers.

Similarly, corporate policies have responsibilities in this regard. Social media platforms should take strict steps against deepfake contents and their creators.

Finally, digital literacy and media literacy should be more available to the common folks. One of the reasons why fake news, deepfakes etc get so much scope to win on the virtual platforms is the public do not have digital literacy, as well as media literacy. They do not see it that way, because there is nothing like individual responsibility on social media. It needs a combined initiative to not let deepfakes win and make the danger real in future.