Deepfakes as a security risk: What's behind it?

28-04-2021

Author: Jan Tissler

"Deepfakes" take deception to a new level: a person's voice and even face can now be automatically imitated. This has consequences for security measures and, not least of all, enables new forms of phishing.

Many of us have already learned not to trust photos. They are too easy to edit and distort. Similar developments are now taking place in the areas of audio and video. One particularly amazing category goes by the name of "deepfake": the face and voice of one person are automatically replaced by those of another, so you can have a person say whatever you want in sound and moving image.

The word deepfake is a combination of deep learning, a method used in the field of artificial intelligence (AI), and fake. This means that instead of painstakingly creating the deception manually, an AI is used to help. Most importantly, the AI needs appropriate material to learn from, for example as many video and audio templates of the target person as possible.

The phenomenon is not entirely new. But the tools for implementing this type of fake are becoming much more powerful. What was more a fun trick that was easy enough to see through a few years ago is now becoming more and more precise. At the same time, the right tools are now available to anybody who's interested. The startup Deeptrace has determined that the number of deepfakes on the Internet increased by 330 percent between October 2019 and June 2020.

And this has consequences for various security measures and technologies.

Example of a deepfake: The fake Tom Cruise

A current example of one particularly amazing deepfake is the TikTok account @deeptomcruise. In a series of short clips, you can see the actor Tom Cruise as he lives and breathes – or at least that's what it looks like. It's hard to believe that these videos are actually fake. You have to look very closely to spot the signs.

The creator behind these viral clips, Belgian video effects specialist Chris Ume, has since explained how they came about. He enlisted the help of a professional Tom Cruise lookalike: Miles Fisher. Firstly, his natural resemblance to the actor provided a good starting point, and secondly, he was able to imitate the facial expressions, gestures and voice.

In the next step, the software was used to modify the face so that Miles Fisher completely morphed into his model Tom Cruise. These days, for simple cases, this last step is done automatically: applications with the right artificial intelligence can analyze two faces and voices, find their special characteristics, and then transform one into the other.

However, the fake Tom Cruise took months to develop and needed a lot of fine-tuning. In this respect, a fake as convincing as this one still takes time and effort to create.

But it still demonstrates just how much is already possible. The technology will be even more advanced in the very near future. And the fake doesn't always need to be that perfect to achieve its purpose.

The dangers of deepfakes

One area in which these fake videos can cause a stir is "fake news" and disinformation campaigns. True, in many cases deepfakes can still be debunked. But we know how quickly sensational news can spread on the web and how stubbornly lies can hold their own against all attempts at clarification.

At the same time, deepfakes can have an impact on security mechanisms. Video identification procedures come to mind. Admittedly, today's technology is not yet advanced enough to generate a sufficiently good fake live. To this extent, technical countermeasures are available. But for how long can we assume that the person on the screen is really the one sitting in front of the camera?

Another possible point of attack is simple facial recognition methods. Apple's Face ID method used in iPhones and iPads is not one of them, given it evaluates not only the camera image, but also relies on other, specialized sensors. A photo or video is therefore not enough to catch someone off guard. It should even be able to recognize masks.

However, not all methods can be considered equally secure if they don't have the necessary hardware. A recent study by South Korea's Sungkyunkwan University, for example, shows that commercially available facial recognition services from providers such as Microsoft and Amazon are vulnerable to deepfake attacks. In some cases, the service even found the fake more convincing than the original.

The good news from the study: existing detection mechanisms for deepfakes worked well in their tests. If they are installed upstream, the services are significantly less vulnerable.

Moreover, the companies concerned are not sitting silently: As part of a "Deepfake Detection Challenge," Amazon, Microsoft, and Facebook collaborated with several universities to research detection methods.

The threat relating to audio is currently more pressing, because voices can already be imitated much better than faces in videos. This means, for example, that new avenues have opened up for phishing attacks, which up until now have primarily used email. And this is not speculation: the manager of a British energy company was apparently tricked into transferring 220,000 euros to a Hungarian supplier. He thought his German superior had asked him to do so in a phone call. The scammers had not only deceptively imitated the voice, but also the typical tone of voice, right down to the accent. And there are other examples too.

Closing words

As explained in another article: The first defense against such attacks is being aware that they are even possible. Phishing often plays on our familiarity with a certain person to weaken our internal alarm systems. And with deepfakes, this no longer affects just emails, but also phone calls and, in the future, even video calls.

In an interview with The Verge, video effects specialist Chris Ume compares it to Photoshop: 20 years ago, only a few people knew which photo manipulations were possible. Today, it's common knowledge. The same thing will happen with deepfakes.

Published by DSwiss AG April 28, 2021