Deepfakes: Harmless Fun or Serious Security Threat?

Machine learning has been utilised in many different ways, but one use we probably didn't see coming is face swap type video filters. Deepfake technology is becoming more accessible and more advanced, but it stops being fun once cybercriminals use it to attack and steal from businesses. So, what are deepfakes, what damage could they do to your organisation, and what does all this have to do with Sea Shanty TikTok?

Deepfakes are back in the news again, but so far, cybercriminals have yet to really deploy spoof AI video on a large scale. Right now, what we'd consider to be deepfake tech is way more commonly used for social media filters, funny videos, or, at worst: fake news and propaganda. However, as the technology becomes more sophisticated, the higher the likelihood that nefarious parties could use it to breach your business.

To combat this, it's good to stay ahead of the curve and get to know what you're up against; familiarise yourself with deepfake technology before that threat becomes a reality. Today's doggy filter could just become tomorrow's major security breach.

What are Deepfakes?

A deepfake is usually a video file (although digital images and audio deepfakes also exist) created using a neural network – a type of machine-learning model, that superimposes the likeness of one person onto the movements of another: almost like a digital mask. The term "deepfake" takes its name from the deep learning technology that's utilised to create these videos. Basically, the AI analyses hours of video footage, learning what a subject's face looks like at all angles, then utilising algorithms to transpose that face onto an actor.

Two complimentary AI systems compete with each other to develop the video's accuracy. To begin with, the generator system creates the fake video clips with the goal of fooling the discriminator system. Every time the discriminator accurately spots a fake, it feeds back to the generator, telling it what not to do. This is called a Generative Adversarial Network. And so, as the generator learns to create more accurate videos, the discriminator learns to spot more forgeries, with the ultimate goal being to create an incredibly sophisticated spoof video that can fool not only the discriminator, but humans too.

Just Fun and Games?

Deepfake style videos find themselves in the news often, most recently due to a viral TikTok video. A quick snap of four friends (posing in their skinny jeans at the beginning of what was no doubt an epic night out) was digitally manipulated to make it look as though they were singing a sea shanty – and it was pretty convincing. This isn't the first

time deepfakes have been used for comedy and entertainment, but there's a darker side to all this…

Politicians are becoming increasingly worried about the threat deepfakes pose to their credibility, and to democracy overall. Deepfake technology has the ability to take fake news to a whole new level, which could negatively impact politics and public trust. Could a political party create deepfakes to smear their opposition? In a world where news outlets rush to report on the most shocking stories, reputations could be destroyed with a click. Things could also swing the other way though; when a trusted public figure is caught out doing something untoward, could they claim the very real footage is actually a deepfake? What will the public's view of video authenticity become when seeing is no longer believing?

How Deepfakes Can Impact Your Business

Thankfully, we're still probably a bit of a way off before cybercriminals can accurately spoof a Zoom call from your boss, but deepfake technology has been used to defraud a company before. In 2019, a deepfake call was used to scam a UK-based energy firm into parting with over £200,000. The company's CEO was targeted by scammers, who posed as the CEO of the energy firm's German parent company. They used AI technology to duplicate his voice, including his subtle German accent and the "melody" of how he spoke. It was enough to convince the UK CEO to transfer funds to what he thought was a Hungarian supplier's account – a costly mistake.

If the deepfake tech was sophisticated enough to fool a CEO in 2019, how many more of us would be fooled by what these AI systems can produce now? Businesses are already being targeted with increasingly sophisticated spear phishing campaigns, but audio and video footage could take things to the next level.

Fighting Back Against Fakes

Tech companies are already working to combat deepfake technology, Microsoft have outlined new tools to combat disinformation, and part of that is being able to recognise deepfakes. Google has released an open source deepfake database and Facebook has offered $10 million to the Deepfake Detection Challenge. On top of that, lawmakers in US passed deepfake legislation as part of the 2020 National Defence Authorisation Act, European police are pushing for something similar in the EU, and hopefully the UK will soon follow suit.

As with most things, knowledge is power when combatting deepfakes, so teaching yourself and your team how to spot them will stand you in good stead for protecting your business and being vigilant about fake news.

Here are a few key giveaways:

- Uncharacteristic behaviour – if your finance director has called you asking to transfer funds to a dodgy account, best to verify that source first.
- Unnatural eye movement/the subject not blinking at all.
- Facial or background glitches when the subject moves suddenly.
- Lack of emotion – if the subject's face doesn't really show the emotion that goes along with what they're supposed to be saying, either they've got an amazing pokerface, or it's a deepfake.
- Awkward body positioning – deepfake tech focuses on replicating the face, so the body is a great giveaway.
- Slightly distorted colouring or features, especially around the forehead and chin.
- Strange audio quality: abnormal or erratic pauses, a slight change in accent or melody.

Just like with suspicious looking emails, if you're sent some audio or video that just feels *off*, always verify it first, especially before you click any links or hand over company cash.

Hear No Evil, See No Evil

Want a more in-depth look at deepfakes? Lucky you, we've done a podcast on this exact subject! Not too long ago, we interviewed Mike Koss, a cybersecurity expert about the threat of deepfake technology.

Check out his insights here:

Hacked Off · 047. Mike Koss: Hear no evil, see no evil

As deepfakes become ever more convincing, the threat they could pose to organisations also increases. We must adapt to this new threat landscape, because verifying our sources – in terms of business, news, and 1830's sea shanties – is more important than ever. Tech that's used for fraudulent purposes is evolving, and we've got to evolve with it to ensure businesses are protected from all angles. What's next for AI generated video? You won't believe your eyes.

If you'd like to know more about protecting your organisation from the latest cybersecurity threats, get in touch with us for a chat, and possibly a sing along too!