The fight against weapons of mass deception

part 1

Albert Barsocchini, Director of Client Advisory Services

Inspire greater. NightOwlDiscovery®

Deepfake is an AI-based technology used to produce or alter video content so that it presents something that didn't, in fact, occur. The word, which applies to both the technologies and the videos created with it, is a portmanteau of deep learning and fake.

The fight against weapons of mass deception

The ability to manipulate audio, video and images has created a fundamental trust problem and is becoming a major corporate threat as well. Soon, we literally might not be able to believe our own eyes or ears when trying to detect and deter corporate fraud.

For example, we may start to see disgruntled employees creating fake clips of their supervisor or co-employee to undermine them or get them fired, or changing a voicemail recording of what a person said to support a false claim. We already see many types of corporate fraud by manipulating email, which is allot easier to detect than so called deepfake video or audio.

Additionally, so called "pump and dump" schemes are using a combination of fake audios and videos to quickly spread corporate misinformation to impact high-frequency trading algorithms that rely on digital information to make investment calls. To fight corporate misinformation, tools are being developed to give digital content a credibility score, so investors can rely on verifiable information about markets.

These threats are magnified by the fact that deepfake apps are becoming ubiquitous and are now easier than ever to download and immediately use to create misinformation or misdirection. To fight back, technologies are being developed by companies like Google and Facebook to automatically check images, video and audio for authenticity.

In the courtroom, it wasn't that long ago that you could assume that if a witness can establish knowledge of the matters depicted in the audio, video or image and can affirm that the proffered exhibit does accurately depict the conditions observed, the exhibit would be accepted into evidence. Sophisticated AI generated images has raised exponentially the degree of difficulty in supporting a legal challenge to digital evidence based on authenticity.

We must learn to question the authenticity of everything that we see and hear and not simply rely on a third party to verify the integrity of the image, since that person may have been misdirected or is part of the deception as well.

This can be challenging as deepfake technologies are becoming more accessible, almost impossible to detect and easier to use. To make matters worse, AI generated media can be trained to outmaneuver forensic technology.

In the legal world of corporate e-discovery, vendors are realizing that they need to have fake busting technologies in their arsenal when required to forensically authenticate the digital media that they are asked to process. Being able to provide an authentication wrapper around digital media may become necessary as the threat becomes more prevalent.

We need to automate the digital authentication process; instead of it being a one-off forensic exercise. Unfortunately, the forensic tools for catching deepfakes are still in its infancy and playing catch up to the growing corporate threat. The arrival of media verification technology may signal the beginning of an AI-powered arms race between audio/video fraudsters on one side and digital media authentication technology on the other.

Being able to provide an authentication wrapper around digital media may become necessary as the threat becomes more prevalent.

For example, [InVid- In Video Veritas] is a browser plugin that can verify video and images circulating on social media. [Foto Forensics] is a website that will run an error level analysis (ELA) to find parts of a picture that were added to it after editing by identifying areas within an image that are at different compression levels. [Findexif.com] is a free tool where you upload a photo or provide a reference to it and Findexif will identify so called EXIF-data (e.g. date and time, device, image characteristics, and GPS location). Amnesty International created [Amnesty YouTube] to verify if a YouTube video has already been posted on the platform before.

The problem with "after the fact" forensic investigations is that often the damage has already been done. Ideally, that means developing technologies for the automated assessment of the integrity of an audio, video or image and that can provide detailed information about how the manipulations were performed.

Part 2 will focus on an actual investigation and best practices.

Are you engaging in media authentication? Unfortunately, seeing (or hearing for that matter) is no longer believing in the digital world. To learn more and understand why it matters contact the author, Albert Barsocchini, directly at abarsocchini@nightowldiscovery.com.

NightOwl Discovery, a global leader in Corporate Discovery Management, helps companies in the most demanding industries reach their governance, discovery, compliance and investigation objectives through a unique portfolio management approach. NightOwl helps enterprise customers maximize investments in people, process, and technology through comprehensive managed service offerings. NightOwl is a data management and advanced analytics company whose offerings span the entire EDRM for customers in the US, EMEA, and APAC. Please contact info@nightowldiscovery.com or visit NightOwlDiscovery.com for more information. **NightOwl Discovery - Inspire greater.**

Inspire greater. NightOwlDiscovery®