SWCP Portal

December 2024

Recognizing Deep Fakes

By Mark Costlow

This month we will talk about strategies to avoid being taken in by AI-generated counterfeits.

Origin of the Term

The term Deep Fake (or deepfake) was first **coined by a Reddit user** with that name in 2017. They created a subreddit for posting pornographic videos with the actors' faces replaced by those of famous people. The "Deep" part of the term comes from "Deep Learning" techniques which use multiple levels of neural networks to swap faces in videos and more.

What started as crude but effective face-swaps has turned into all-encompassing media fakery. AI technology is making it easy to swap faces, change voices, match new words to existing video, and even create whole apparently-human characters out of thin air.

Imitation media can come in any form: still images, voices, or video. The reasons to make fakes range from beneficial to benign to malicious. We might enjoy when Hollywood does it, but it can also be used for theft and dangerous manipulation.

Examples of Deep Fakes

One classic example of deep fake video is **Mark Zuckerberg in 2019** (apparently) saying in an interview with CBS News that he is happy to have stolen the data of billions of users and is using it to manipulate them.

Another example, from 2020, is the **Belgian Premier delivering a speech** claiming that the climate crisis is responsible for Covid-19, SARS, and Ebola. The fake was made by a group of climate activists trying to keep focus on climate issues in the midst of pandemic lock-downs. There might be links between climate change and pandemics, but the Belgian Premier never said anything like that. The fake video co-opted her likeness for the activist's purposes.

More recently, **CNBC reported** that primary voters in New Hampshire heard a deepfake of President Biden asking them **not** to vote in the presidential primary. **CNN reported** that a finance worker sent a $25 Million payment after a video call with a deepfaked version of his CFO.

AI technology has advanced by leaps and bounds in the last five years, and these early examples might seem crude by today's standards. Any media you encounter may have been modified, or completely created, by AI. Sometimes it is easy to spot the fakes, but as the fabrications get more and more sophisticated, identifying them gets correspondingly harder.

Positive Uses?

AI deepfakes are a tool, and tools are not inherently good or evil. The people who wield them determine that. The overwhelming uses of deepfakes appear to be sinister, but sometimes they are used for good.

Filmmakers and video game producers specialize in bringing to life ideas and images that never occurred in the real world. Millions of us enjoy the computer effects that allow Tom Cruise to swap faces with bad guys in movies, or bring to life ten-foot-tall blue-skinned Na'vi in the movie Avatar.

Even this beneficial use has a dark side though. A key point of contention in the recent actors' and writers' **strikes centered around AI-generated content**. Actors worry that motion capture plus deepfake technologies will let studios re-use an actor's likeness in new ways without compensation.

Another favorable use is disguising your identity online with a voice-changer. **Voicemod** and others make "voice skin" software that modifies your voice in real-time to sound like a fictional character or even a known actual person. There are some **good reasons to want to to this**, such as avoiding a bully, separating your work-life and non-work-life, or disguising your age or gender to avoid stereotypes. In extreme cases it is used to thwart stalkers or make anonymous reports of crimes.

The common dark side of voice fakes is in scams. A caller can impersonate a family member in trouble as a way to con a scared relative out of thousands. In a hilarious double-reversal, **Youtuber Kitboga** uses a voice skin to sound like an elderly grandmother. He engages scammers in long-running anti-scams, eating up their time. His goal is to thwart them by getting their accounts removed, or at the very least, waste days of their time that would otherwise be used to victimize others.

How to Spot Them

You know those **counterfeit banknote detection pens** they use at retail counters to spot fake bills quickly and reliably? Unfortunately we don't have anything like that for AI deepfakes, and as technology improves, techniques that work today might not work as well tomorrow. For now, here are some ways to spot the counterfeits:

1. Look for Subtle Inconsistencies

Deepfakes often have slight glitches or unnatural face or body movements. Often the the audio dialogue might not be perfectly synchronized with the mouth shape and movement. Our brains have had eons of training in how a typical human moves and acts, so we are pretty good detectors of fakes when we slow down and pay careful attention.

2. Check the Lighting and Shadows

Look for inconsistencies between the scene lighting and the apparent light falling on the objects. Shadows on the wrong side or completely missing are dead giveaways. For pictures of people outside, shadows of humans, their hair, and whatever they are carrying are good places to spot problems.

3. Audio Clues

In addition to watching for mismatches between the audio and the speaker’s facial expressions, listen for robotic or unusual speech patterns. Voice synthesis has gotten very good, but sometimes fakers let mistakes slip through.

4. Watch the Eyes

Many deepfakes struggle to create natural eye movements or maintain a consistent gaze direction.

5. Unstable Backgrounds

Backgrounds might seem to warp or shift strangely, especially around the edges of the manipulated person.

6. Use Deepfake Detection Tools

VLink has a list of popular Deepfake detection tools. However, this answer is almost a deepfake in itself. The tech press has many stories about deepfake detection software, including Microsoft's Video Authenticator and Intel’s FakeCatcher. The latter even uses "blood flow" analysis!

But when you try to find or use these tools, they dissolve into wisps of vaporware. There are real tools out there, but they are “projects” that require sophisticated use of cutting-edge tools, not packaged utilities available in an app store. Maybe somebody knows of a good easy-to-use tool and will write in to tell us about it!

7. Look for Additional Sources

If the video supposedly happened in public, it is likely to have been captured by other cameras at the same time. Between smartphones, CCTV, dash cams, and body cameras, it is becoming unusual for something to be captured only once. Did they all see the same thing?

8. Verify the Source

Your own skepticism may be the number one tool for detecting AI fakes. Deepfakes are just another form of a scam, and like all scams they are greatly helped by our own inattention. If someone says something that makes you think, "I can't believe they said that!" then slow down, consider whether they really would say that, and then look for independent verification.

Beware in this process not to accept the word of outlets that are just parroting the original social media post without themselves verifying its veracity. Forwarded emails with wild claims are notoriously rife with made-up stories.

Test Yourself

Northwestern University's Kellogg School of Management created a **Detect Fakes experiment** anyone can try.

The quiz format offers a few hundred images to evaluate. After you've done at least 5, it tells you how you are doing compared to past participants. I scored slightly higher than the average with 80% success. But, I was paying close attention and trying hard. If I encountered those same 10 images randomly online, my score would be much lower.

15 Minutes Into The Future

Andy Warhol is often **credited with saying**, “In the future, everyone will be world-famous for 15 minutes.” It turns out it wasn’t said by Warhol, but rather appeared in a brochure about his 1968 exhibition in Stockholm, Sweden.

Whoever first thought up the phrase, it certainly seems to have come true. Old media manifested the phenomenon to a degree, giving us people who are “famous for being famous”. Social media really revved it up though. Every month there’s a new name on everyone’s lips that we’ve never heard of before. They probably haven’t done anything truly interesting, just went viral in an accident of circumstance (Pizza Rat for examples).

I will close out the year with a related phrase I heard recently. It came from **Jack Rhysider on the Darknet Diaries podcast** and it somehow sums up our fears about the future of our globally interconnected world. It made me laugh, then gave me chills, then made me laugh again (so as not to weep).

> In the future, everyone will
> have 15 minutes of privacy

If there is one concept I would like people to remember from the past year of this newsletter, it is that your own mind is your best defense against those who would mislead you. Slow down, think it through, don’t react until you’ve considered the situation. See you in 2025!

Southwest Cyberport - swcp.com
New Mexico’s Expert Internet Service Provider since 1994
505-232-7992 | Support: help@swcp.com
5021 Indian School NE, Ste. 600, Albuquerque, NM 87110
Click on bold blue type in browser for links.