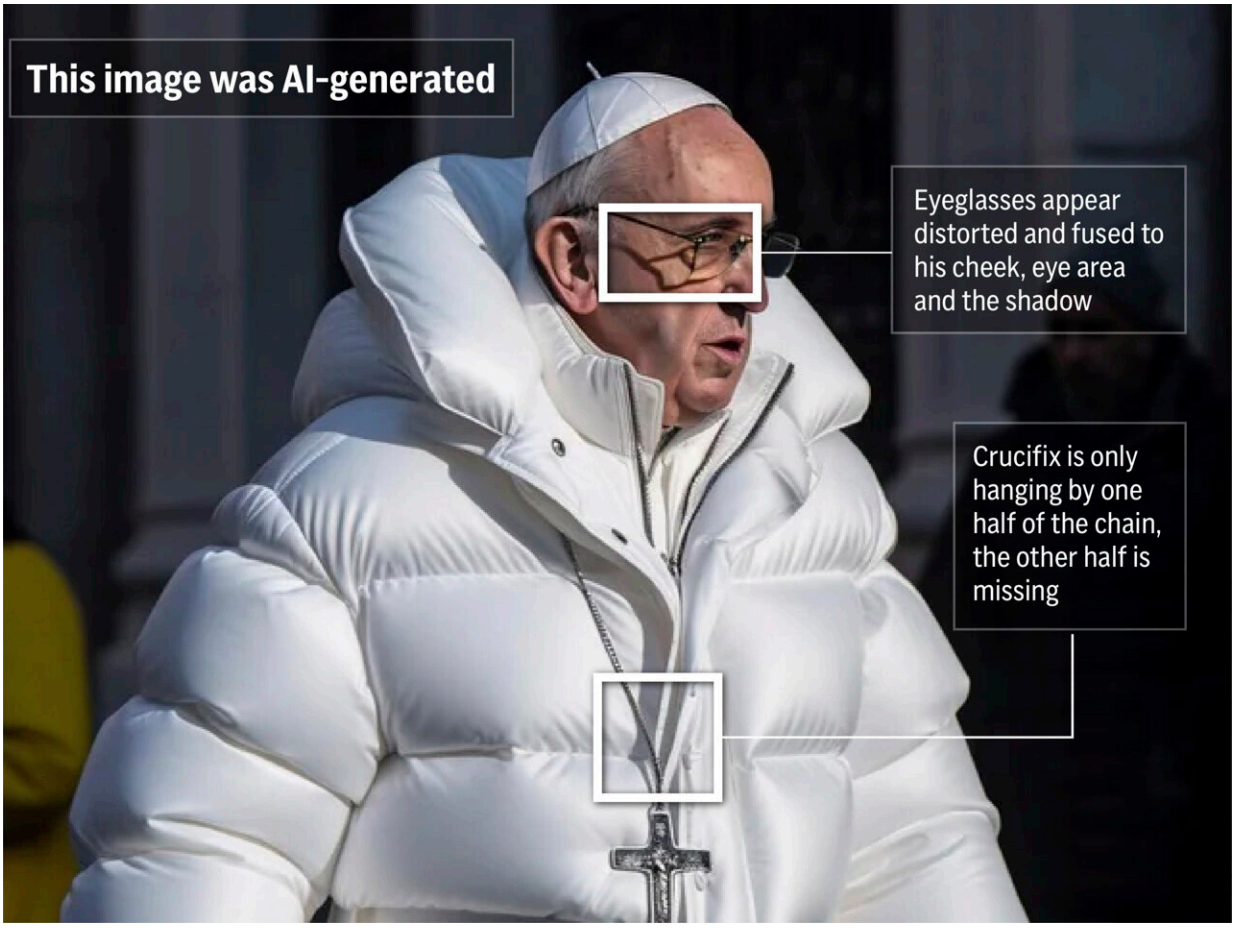
One Tech Tip: How to spot AI-generated deepfake images

March 21 2024, by KELVIN CHAN and ALI SWENSON



AI fakery is quickly becoming one of the biggest problems confronting us online. With AI deepfakes cropping up almost every day, depicting everyone from Taylor Swift to Donald Trump, it's getting harder to tell what's real from what's not. The following photo-illustrated graphic highlights a few notable areas of an AI-deepfake of Pope Francis.

AI fakery is quickly becoming one of the biggest problems confronting us online. Deceptive pictures, videos and audio are proliferating as a result of the rise and misuse of generative artificial intelligence tools.

With AI deepfakes cropping up almost every day, depicting everyone from Taylor Swift to Donald Trump, it's getting harder to tell what's real from what's not. Video and image generators like DALL-E, Midjourney and OpenAI's Sora make it easy for people without any technical skills to create deepfakes—just type a request and the system spits it out.

These fake images might seem harmless. But they can be used to carry out scams and identity theft or propaganda and election manipulation.

Here is how to avoid being duped by deepfakes:

HOW TO SPOT A DEEPFAKE

In the early days of deepfakes, the technology was far from perfect and often left telltale signs of manipulation. Fact-checkers have pointed out images with obvious errors, like hands with six fingers or eyeglasses that have differently shaped lenses.

But as AI has improved, it has become a lot harder. Some widely shared advice—such as looking for unnatural blinking patterns among people in deepfake videos—no longer holds, said Henry Ajder, founder of consulting firm Latent Space Advisory and a leading expert in generative AI.

Still, there are some things to look for, he said.

A lot of AI deepfake photos, especially of people, have an electronic sheen to them, "an aesthetic sort of smoothing effect" that leaves skin "looking incredibly polished," Ajder said.

He warned, however, that creative prompting can sometimes eliminate this and many other signs of AI manipulation.

Check the consistency of shadows and lighting. Often the subject is in clear focus and appears convincingly lifelike but elements in the backdrop might not be so realistic or polished.

LOOK AT THE FACES

Face-swapping is one of the most common deepfake methods. Experts advise looking closely at the edges of the face. Does the facial skin tone match the rest of the head or the body? Are the edges of the face sharp or blurry?

If you suspect video of a person speaking has been doctored, look at their mouth. Do their lip movements match the audio perfectly?

Ajder suggests looking at the teeth. Are they clear, or are they blurry and somehow not consistent with how they look in real life?

Cybersecurity company Norton says algorithms might not be sophisticated enough yet to generate individual teeth, so a lack of outlines for individual teeth could be a clue.

THINK ABOUT THE BIGGER PICTURE

Sometimes the context matters. Take a beat to consider whether what you're seeing is plausible.

The Poynter journalism website advises that if you see a public figure doing something that seems "exaggerated, unrealistic or not in character," it could be a deepfake.

For example, would the pope really be wearing a luxury puffer jacket, as depicted by a notorious fake photo? If he did, wouldn't there be additional photos or videos published by legitimate sources?

USING AI TO FIND THE FAKES

Another approach is to use AI to fight AI.

Microsoft has developed an authenticator tool that can analyze photos or videos to give a confidence score on whether it's been manipulated. Chipmaker Intel's FakeCatcher uses algorithms to analyze an image's pixels to determine if it's real or fake.

There are tools online that promise to sniff out fakes if you upload a file or paste a link to the suspicious material. But some, like Microsoft's authenticator, are only available to selected partners and not the public. That's because researchers don't want to tip off bad actors and give them a bigger edge in the deepfake arms race.

Open access to detection tools could also give people the impression they are "godlike technologies that can outsource the critical thinking for us" when instead we need to be aware of their limitations, Ajder said.

THE HURDLES TO FINDING FAKES

All this being said, artificial intelligence has been advancing with breakneck speed and AI models are being trained on internet data to produce increasingly higher-quality content with fewer flaws.

That means there's no guarantee this advice will still be valid even a year from now.

Experts say it might even be dangerous to put the burden on ordinary people to become digital Sherlocks because it could give them a false sense of confidence as it becomes increasingly difficult, even for trained eyes, to spot deepfakes.