

Your brain is better at busting deepfakes than you are

July 11 2022, by Loren Smith



Deepfake faces used in the study. Credit: Moshel et al.

While observers can't consciously recognize the difference between real and fake faces, their brains can, University of Sydney research shows.

When it comes to deepfakes, what if fact could be distinguished from fraud? University of Sydney neuroscientists have discovered a new way:

they have found that people's brains can detect AI-generated fake faces, even though people could not report which faces were real and which were fake.

Deepfake videos, images, audio, or text appear to be authentic, but in fact are computer-generated clones designed to mislead you and sway public opinion. They are the new foot soldiers in the spread of disinformation and are rife—they appear in political, cybersecurity, counterfeiting, and border control realms.

For example, in 2016, a Russian troll farm deployed over 50,000 bots on Twitter, using deepfakes as profile pictures, to try to influence the outcome of the U.S. presidential election. Some research suggests they may have boosted Donald Trump's votes by more than 3%. More recently, a [deepfake](#) video of Volodymyr Zelensky urging his troops to surrender to Russian forces surfaced on Twitter, Facebook, and YouTube, muddying people's understanding of the war in Ukraine with potential, high-stakes implications.

When looking at participants' brain activity, the University of Sydney researchers found deepfakes could be identified 54% of the time. However, when participants were asked to verbally identify the deepfakes, they could only do this 37% of the time.

"Although the brain accuracy rate in this study is low—54%—it is statistically reliable," said senior researcher Associate Professor Thomas Carlson, from the University of Sydney's School of Psychology.

"That tells us the brain can spot the difference between deepfakes and authentic images."

Spotting bots and scams

The researchers say their findings may provide a springboard in the fight against deepfakes.

"The fact that the brain can detect deepfakes means current deepfakes are flawed," Associate Professor Carlson said. "If we can learn how the brain spots deepfakes, we could use this information to create algorithms to flag potential deepfakes on digital platforms like Facebook and Twitter."



Can you tell real from deepfake? The image on the left is real. Credit: Moshel et al.

Further into the future, they project that technology, based on their and similar studies, could be created to alert people to deepfake scams in real time. For example, [security personnel](#) might wear EEG-enabled helmets that could alert them that they may be confronting a deepfake.

Electroencephalography (EEG), a test that reveals activity of the surface layer of the brain, was used in the researchers' study.

Associate Professor Carlson said: "EEG-enabled helmets could have been helpful in preventing recent bank heist and corporate fraud cases in Dubai and the U.K., where scammers used cloned voice technology to steal tens of millions of dollars. In these cases, finance personnel thought they heard the voice of a trusted client or associate and were duped into transferring funds."

Method: Eyes versus brain

The researchers performed two experiments: one behavioral and one using neuroimaging. In the behavioral experiment, participants were shown 50 images of real and computer-generated fake faces. They were asked to identify which were real and which were fake.

Then, a different group of participants were shown the same images while their [brain activity](#) was recorded using EEG, without knowing that half the images were fakes.

The researchers then compared the results of the two experiments, finding people's brains were better at detecting deepfakes than their eyes.

A starting point

The researchers caution that given the novelty of this field of research, their study, published in *Vision Research*, is just a starting point. It won't immediately—or potentially ever—lead to a foolproof way of detecting deepfakes.

Associate Professor Carlson said: "More research must be done. What gives us hope is that deepfakes are created by computer programs, and these programs leave 'fingerprints' that can be detected.

"Our finding about the [brain](#)'s deepfake-spotting power means we might have another tool to fight back against deepfakes and the spread of disinformation."

More information: Michael L. Moshel et al, Are you for real? Decoding realistic AI-generated faces from neural activity, *Vision Research* (2022). [DOI: 10.1016/j.visres.2022.108079](https://doi.org/10.1016/j.visres.2022.108079)

Provided by University of Sydney

Citation: Your brain is better at busting deepfakes than you are (2022, July 11) retrieved 6 May 2024 from <https://medicalxpress.com/news/2022-07-brain-deepfakes.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--