'Is this a deepfake?' Why we're asking the wrong question

April 8 2024, by Justin Zadorsky

Aleksander Essex created this deepfaked image of the Mona Lisa as an easily spotted altered image. Credit: Aleksander Essex

Over the past year, the prevalence and potential dangers of deepfakes have raised concerns related to personal privacy, business ethics and even election interference.

As a result, advice about spotting deepfakes has been circulating on social media and in the news, but engineering professor Aleksander

Essex says we should start thinking beyond deepfake detection and start asking ourselves more important questions.

Essex runs Whisper Lab, the Western Information Security and Privacy Research Laboratory, and spoke with Western News about deepfakes and what they mean for our future society.

What is a deepfake?

I don't think we've settled on a single definition yet, but we're talking about the use of deep learning techniques (a type of AI) to apply the likeness of real people to fictional imagery, audio or video.

Are deepfakes problematic?

Definitely. We have plenty of real-world examples, like when a financial worker paid out $25 million (USD) following a call with a deep-faked chief financial officer. There was news about a website used for generating fake photo IDs, and earlier this year, Twitter suspended searches for Taylor Swift after deepfake nude imagery flooded the platform.

In your work around election security, are deepfakes becoming an issue?

Fortunately, it hasn't come to Canada in full force yet. But we can expect it will eventually, as it's already happening in the U.S. We have seen several high-profile examples, such as a deepfake Joe Biden robocalling New Hampshire voters. Fortunately, the federal government recently tabled new legislation that targets false statements around election practices, which covers deepfakes.

What are some ways to spot deepfakes?

In the near term, you can still often trust your instincts about deepfakes. The mouth moves out of sync with the body, or reflections are at a different frame rate, etc.

In the medium term, we can use deepfake detection software, but it's an arms race, and the accuracy will likely decline over time as deepfake algorithms improve.

In the long term, deepfakes may eventually become indistinguishable from real imagery. When that day comes, we can no longer rely on detection as a strategy. So, what do we have left that AI cannot deepfake? Here are two things: physical reality itself and strong cryptography, which is about strongly and verifiably connecting data to a digital identity.

Cryptography is what we use to keep browsing histories private, passwords secret, and it lets you prove you're you. The modern internet could not exist without it. In the world of computation, AI is just an algorithm like all others and cryptography is designed to be hard for any algorithm to break.

We are still able to link a physical entity (a person) to a strong notion of digital identity. This suggests that 'is this a deepfake?' may not be the right question we should be asking.

If 'is this a deepfake' is the wrong question, what is the right one?

The right questions to ask are: Where is this image coming from? Who is the source? How can I tell?

The sophistication of deepfakes may eventually evolve to the point where we can no longer distinguish between a real photo and an algorithmically generated fantasy.

In this world, the focus of the conversation should be less on the content of the image but on where it came from, i.e., the source, the communication channel, the medium. In that sense, Marshall McLuhan's old wisdom that "the medium is the message" still applies—perhaps now more than ever.

What would you say to those worried about the potential dangers of AI and deepfakes?

I spend a lot of time talking to my students about the dangers of magical thinking when it comes to new technologies. AI, at least the kind that runs on a normal computer, is not alive or conscious. Yes, it can be useful. Maybe even harmful. But it's just gears spinning in an elaborate cuckoo clock and society needs not to run away with itself.

In the end, there are fundamental limits to computation, just like the speed of light. AI cannot do everything. Nor can quantum computing or blockchains. Just remember: AI is just a tool. Our attention should be on the hands that wield it.

Provided by University of Western Ontario