



# Unmasking Deepfakes

## Navigating the Shadows of Digital Deception

Author: **James Moffett**

**THINK** explores how aspects of deepfake technology work and investigates the risks that come with it.

**D**espite the widespread use of generative AI such as ChatGPT, Gemini and DALL-E, many users are still oblivious to the inner workings and sophisticated computing that lie behind it. **THINK** sat down with Dr Dylan Seychell, lecturer at the Department of Artificial Intelligence at UM, to clear the air and discuss how generative AI actually works.

Simply put, generative technologies refer to a category of tools and systems that use artificial intelligence to generate new content, ideas, or approximate solutions. Rather than following pre-programmed instructions, these technologies can sift through existing data and learn specific patterns to produce unique results.

To improve the outcomes of such technologies, training is required. AI Systems such as DALL-E, ChatGPT, and Gemini are being taught with

a colossal amount of data sets – a collection of data to test and train algorithms and models. For instance, a few years back, AI models were being trained on tens of thousands of images to identify particular items or objects. Nowadays, the sophisticated nature of these models requires it to train on billions of images. 'Once you are processing all that information, you are beginning to create relationships with different images that, as human beings, we are not capable of creating; and this is something which helps us work together with AI,' Seychell explains.

### THE PITFALLS AND BENEFITS OF DEEFAKE TECHNOLOGY

The sophisticated nature of AI has granted users scenarios which could not be conceived before. Tools are being designed in such a way that allows anyone to create photos of someone by inputting descriptions of a person – enabling the technology to then create an image based on

the details provided. 'When we talk about deepfakes, besides giving some form of instruction to the AI tool to create a photograph, we can also give visual descriptions and images of a real person.' Therefore, the generated content will be created using the visual characteristics of the real individual being described. In this way, one can use someone's photograph and place that person in another image or video, and in a situation where this person was not actually present.

'There are a lot of dangers of using deepfake technology. I struggle to find any positives in it, to be honest,' maintains Seychell. 'It rarely happens, but when I weigh this particular technology, it is difficult to find anything advantageous about it.' One of the ways deepfakes work is through what is referred to as Generative Adversarial Networks (GANs).

To better understand GANs, let's imagine a scenario: a money forger and the police who are trying to find out



which money is a forgery. The forger needs to produce a false banknote, so he uses a napkin and writes down €500 on it with a pen. He hands it to the police, who realise it is not real money. So the forger then prints a piece of paper on a computer with €500 typed on it. He tries again, but the police still identify it as a fake. The forger then adds a bit of colour and adjusts the banknote's size correctly, and the police start to doubt but still realise it is fake money. Eventually, the forger manages to produce a banknote where the police are unable to distinguish between it and a real one.

Similarly, this process creates sophisticated and learned AI algorithms, which have been designed in such a way as to teach and learn from each other, via the generator (the forger) and the detector (the police). Despite the benefits of having an AI model teach itself in such a way, issues may soon arise that can spiral out of control. 'If the detector being placed is more sophisticated in detecting fakes, the generator will keep on generating even more realistic and convincing content,' explains Seychell. This will happen in turn until the detector fails to detect what is real or fake. 'It learns and keeps on improving,' continues Seychell. 'The more we improve detection, the more we are improving the generated

*2024 is a record for elections worldwide, with around 40 to 60 in total. This year we will witness the real test of how these things escalate.*

content.' In this sense, using GANs makes the situation worse when it comes to recognising deepfake content.

Nowadays, countless online scams disguised as advertisements have infiltrated every nook and cranny of cyberspace. Using deepfake technology to place well-known individuals in such ads and asking viewers to invest in fraudulent business schemes has become the norm. *TIME Magazine* points out that 2024 is a record for elections worldwide, with around 40 to 60 in total. This year we will witness the real test of how these things escalate.

While Seychell believes that deepfake technology does not provide many positive outcomes, there are certain advantages of employing it in particular fields. Two such possible

instances are the cultural and medical sectors. Within the cultural sphere, visitors to an interactive centre may have the possibility of speaking to historical figures – brought back to life, as it were, through AI's sophisticated approach – which can instil a deeper appreciation for history in general. In addition, techniques involving neural networks and architecture can also be used to generate synthetic data that allows the training of algorithmic models to predict heart diseases or approximate the production of certain new medicines.

'There are certainly positives from this type of technology, but from the perspective of vision – when you see it generating something visual – that is where it starts to be difficult to find the pros. When you combine this technology with society, it has some ugly implications,' says Seychell. As with anything else, technologies are sometimes twisted to be used for nefarious purposes, despite their innocent origins.

## **BUILDING A SENSE OF INTUITION**

Despite persistent warnings to be on the lookout for false advertisements, fake scam calls, and the like, with the ever-increasing sophistication of deepfake algorithms, identifying



what is real or not is becoming more problematic. As an academic and researcher on artificial intelligence, Seychell offers some pointers one can follow when encountering potentially misleading information. 'When we come across something that is not normal, or we are not completely expecting it, let us find alternative ways to verify it.' As technology users, we need to build a sense of intuition so that we do not jump to conclusions. Especially when it comes to money, Seychell's rule of thumb is to always check first. 'If we think someone we know is asking for money, but they do not normally do, pick up the phone and call them.' Verifying the legitimacy of the information before acting is a sure way to avoid falling for such deceptions.

Using this financial scenario as an example, Seychell's take on this issue extends further – outlining the clear-cut distinction between what constitutes technology and what makes us humans. 'If someone I know is asking for money, besides verifying it, as a human being I should care enough to check up on that individual to see how I can help them directly.' We need to look at ourselves and examine how we use technology in addition to the social aspect of us as human beings. What we need to ask is whether it is

more convenient to send money rather than care for the person, and seeing how many are victimised by such scams, could this imply that we may be losing touch with our human selves?

### TECHNOLOGY AS OPPORTUNITY

Despite the pitfalls and concerns that are brought about with every new technological development, as Seychell sums up, 'let us also not be afraid of technology. Even though we discussed the dangers, deepfakes happen to be the only topic that worries me.' Yet, there is a wealth of advantages and positive outcomes from the use of

*'If someone I know is asking for money, besides verifying it, as a human being I should care enough to check up on that individual to see how I can help them directly.'*

AI technology in general that we are now becoming more dependent upon. Without its use, human beings would regress, given how ingrained within our societies artificial intelligence has become – and is continuing to do so. Being aware of the utility of these tools, along with their possible negative applications, will arm users with enough knowledge to proceed at a steady, if cautious, step towards the next stage in our own evolution. Ultimately, technology aims to transform challenges into opportunities. **T**

#### Further Reading

Ewe, K. (2023). *The Ultimate Election Year: All the Elections Around the World in 2024*. *TIME*. <https://time.com/6550920/world-elections-2024/>

Thambawita, V., Isaksen, J. L., Hicks, S. A., Ghouse, J., Ahlberg, G., Linneberg, A., Grarup, N., Ellervik, C., Olesen, M. S., Hansen, T., Graff, C., Holstein-Rathlou, N.-H., Strümke, I., Hammer, H. L., Maleckar, M. M., Halvorsen, P., Riegler, M. A., & Kanters, J. K. (2021). Deepfake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-01295-2>