Q&A: What are deepfakes and should we be worried?

May 22 2024, by Candy Gibson

Credit: AI-generated image

Deepfakes are creating havoc across the globe, spreading fake news and pornography, being used to steal identities, exploiting celebrities, scamming ordinary people and even influencing elections.

Yet a worldwide survey found 71% of people have no idea what

deepfakes are.

Deepfakes are digital photos, videos or voices of real people that have either been synthetically created or manipulated using artificial intelligence (AI) and can be hard to distinguish from the real thing.

You've probably seen a deepfake video or photo without even realizing it. Computer-generated Tom Cruises, Taylor Swifts and Mark Zuckerbergs have been circulating on the internet for several years, but what started as a bit of harmless fun has now become much more serious.

Earlier this year, a finance worker at a multinational firm in Hong Kong was tricked into paying AUD$39 million to fraudsters who used deepfake technology to impersonate the company's chief financial officer in a video conference.

In 2022, a fake video of Ukrainian president Volodymyr Zelenskyy emerged, falsely portraying him urging his military to surrender to invading Russian forces. While this was quickly shut down by the Ukrainian leader, there are real fears that deepfakes will spread false information and conspiracy theories in multiple election campaigns this year, including in the US, UK, India and Russia.

Australia's Defense Chief Angus Campbell has expressed fears that the world is entering "an era of truth decay," where misinformation will undermine democracy by sowing discord and distrust.

Campbell told a defense conference in 2023 that artificial intelligence and deepfakes would "seriously damage public confidence in elected officials" by making it impossible for most people to distinguish fact from fiction.

For some years, computer scientists—including those in UniSA STEM—have been developing novel AI technologies to help answer important challenges in industry, health care, engineering and defense.

Through advances in machine learning and deep neural networks, they have used the technology for good—but there has always been the potential for people with malicious intent (known as "bad actors" in the industry) to turn it to their advantage. Enter deepfakes.

In this feature, Associate Professor Wolfgang Mayer and Professor Javaan Chahl provide their perspective on deepfakes.

Associate Professor Wolfgang Mayer, UniSA computer scientist and AI expert

Q: Deepfakes are possible through advances in machine learning and AI. What positive gains have we made due to this technology, and does this counter the dangerous uses?

Generative AI technology has always had good and bad uses.

We hear a lot in the media about how generative AI is being used for harmful purposes, but it is doing more good in the world than bad. If you think about self-driving cars to avoid accidents and allow disabled people to move around freely, that is only possible through artificial intelligence and machine learning.

This technology allows us to accelerate medical health research and detect diseases earlier; it is also a powerful tool in construction and engineering, freeing up mundane tasks and minimizing errors; and it is the foundation for computer visions systems. The positive uses far outweigh the negatives.

However, there is no doubt that deepfakes are causing issues. The problem of propaganda and misinformation is not new, but it is now on a different scale due to deepfakes.

Q: Most people would assume that AI is only understood by computer scientists, specialized IT engineers and not ordinary people. True?

It's a complicated process to build systems inspired by the brain, but to use these systems is now relatively easy—and that's why deepfakes are proliferating. The systems have become powerful enough that we don't need to be machine learning experts to use them. It's just a matter of downloading an app that is developed by tech companies.

Q: Will putting a digital watermark on authentic AI images help address the proliferation of deepfakes?

No. It might stop the most simplistic users, but the serious ones will be able to replicate the technology without watermarking it. As the quality of generative AI improves, it will be more difficult to detect fake photos, videos and cloned voices.

Q: What are some of the ways you can spot deepfakes?

It can be tricky, depending on how much effort someone has put in to creating a deepfake. Synthesizing hands is always difficult and often the lips do not synchronize with the voice in a video. However, replicating faces in a video conference is much easier. I think ultimately it will be extremely hard to spot deepfakes based on appearance. We need to rely on what is being said, what the situation is and whether it's odd in some way.

Q: What can people do to protect themselves from being the target of deepfakes?

Unless you are in a position of power, or a celebrity, you are probably not at risk of being copied. However, people need to be careful what they put online because all that data can be used to mimic us.

If I were to put all my lecture notes online it wouldn't be that difficult to generate a virtual Wolfgang to give my lectures, but fortunately they are behind a paywall. It is a good idea to verify from other sources whether what you have seen or heard—especially on social media—is in fact accurate.

Professor Javaan Chahl, DST/UniSA Joint Chair of Sensor Systems

Q: What impact will deepfakes have on governments and could they erode trust and confidence in our leaders?

Governments and politicians are already undermining democracy. At the fringes of every election campaign, particularly in the past decade, there have been fake flyers; things that are said by politicians that turn out later to be egregiously false. That game is already well afoot.

To mitigate a sense of panic around deepfakes, misinformation from higher powers has been going on for an awfully long time. The reason senior levels of government might be worried about this use of AI is because they have had a stranglehold on information in the past and they could decide what you get told. Now, faceless people can start throwing noisy signals into the system and cause chaos.

In Australia, our relationship with power is distrust and always has been. I don't think people trust leadership as much as leaders think they do, and the infiltration of deepfakes means they will have to work that much harder to assure the public they are telling the truth. That's not a bad

thing.

Q: What role do the media have in ensuring they do not disseminate deepfakes?

News outlets are already using ChatGPT to write their stories, so don't be surprised if people stop trusting them altogether. Viewers should be questioning every video they see on social media and mainstream news channels in any case, because they usually only include segments that suit their narrative.

News outlets have become more partisan in recent years and that is eroding the public's trust even without the infiltration of deepfakes. The media need to pursue a non-partisan stance, rely more on facts, and start verifying where media comes from if they want to be trusted.

Q: Are we entering an era where it is hard to separate fact from fiction due to AI?

People not automatically believing what they see, read or hear from digital media is probably a good thing. We need more critical thinking rather than accepting everything at face value. My advice is to question anything that you see on social media or the news unless you know the person who has filmed it and you trust them implicitly.

Even images and video that have not been digitally manipulated can contain falsehoods in the form of propaganda, so people should treat all videos like movies, which are essentially an exercise in creating an artificial world that resembles the real world to varying degrees.

Q: Should we be worried about deepfakes and AI infiltrating our defense force and posing a threat to our national security?

Misinformation has been around for decades—it is just in a different form now. We are part of it. In the past, rival regimes exploited personality flaws common among intellectuals to export clever, seductive and divisive ideology. These destabilization operations are still underway long after those regimes are gone.

Now we have very sophisticated technology that is being deployed in the form of cyber warfare, disrupting vital computer systems for strategic or military purposes, and deepfake technology to diminish social cohesion. It is all about the manipulation of information, which has always been part and parcel of conflict.

Q: What can researchers like yourself do to address deepfakes, or are we just fighting a losing battle?

Right now, we can use computer vision technology to read vital signs on a video to see whether it is fake. We know that deepfake videos have irregular vital signs beyond breathing and heart rates—such as blood oxygen saturation, blood pressure, temperature and other things that can indicate whether it is fake. That might buy us a few years, but eventually the videos are only as good as the human chain of custody that led to the video.

Provided by University of South Australia

Citation: Q&A: What are deepfakes and should we be worried? (2024, May 22) retrieved 16 June 2024 from https://techxplore.com/news/2024-05-qa-deepfakes.html