Read to be SURE

Volume 4—2024/25

s.u.r.e.
Source · Understand · Research · Evaluate

Issue 1

DEEPFAKES

All information and weblinks are valid as of March 2025.

Deepfakes are BAD BUT can its technology be used for GOOD?

Deepfakes

Deepfakes, created by artificial intelligence, imitate a person's appearance or voice with remarkable accuracy. Inauthentic content has become easier to produce, and at low to no cost for its creator. As technology advances, deepfakes are increasingly becoming indistinguishable from authentic content.

Singapore is experiencing a rise in deepfake cases, mirroring a global trend of rapidly increasing AI-generated fraudulent content. Experts advise the public to remain vigilant, warning that such content is likely to become increasingly sophisticated. According to a 2024 survey conducted by Jumio, a digital identity verification company, 88 per cent of Singaporeans worry daily about falling victim to deepfake scams. However, 69 per cent of Singaporeans trust the government's ability to effectively regulate AI.

While AI technology offers some promising applications such as personalised support for students and digital avatars in films, deepfakes have led to significant negative consequences. These include the spread of misinformation and disinformation, identity theft for fraudulent activities, and the erosion of trust in online platforms and media. AI-generated disinformation also has the potential to manipulate public opinion and influence election outcomes.

In response to these growing threats, Singapore is developing tools, regulations, and public awareness to tackle the challenges posed by deepfakes. These efforts include a $20 million initiative to create detection tools and the establishment of the Centre for Advanced Technologies in Online Safety in 2024, aimed at enhancing expertise in combating deepfakes and misinformation. Singapore also set up a global open-source community to bring together AI stakeholders and policymakers in building trustworthy AI. Additionally, the Online Criminal Harms Act allows the government to issue directions and orders to online platforms to prevent potential scams, including deepfake-enabled content, from reaching Singapore users. The Smart Nation 2.0 initiative also aims to use innovative learning approaches to improve Singaporeans' resistance to deepfake-enabled manipulation, while introducing a new Code of Practice to require specified social media service providers to prevent and counter abuse of digitally manipulated content on their services.

When encountering content that seems genuine but raises doubts, it is important to fact-check the information, find the original source, and rely on credible evidence from authoritative sources.

So, deepfakes are bad, but can the technology be used for good? Let's hear what the word on the street is:

READ TO BE SURE
Read to be SURE Hits the Streets:
Deepfakes and AI Technology

So, does this technology have more promising opportunities or pose major risks to society?

AI technology is beneficial to us.

AI technology holds beneficial uses and potential for good.

AI technology and AI-generated media could contribute to various fields such as entertainment, education, business, advertising, art, and culture (e.g. museum visitors taking selfies with Dali).

Generative AI tools have also been utilised in industry and research, particularly in clinical research and healthcare. AI-generated characters could enhance health and well-being by providing personalised guidance and therapeutic interactions.

For example, the Deep Empathy project uses AI-generated visuals to foster empathy for disaster victims by simulating how cities might look if they faced conflicts like those in war-torn Syrian neighbourhoods.

New measures are in place to protect the digital landscape from deepfakes.

Technological solutions such as social media platform scanning tools are crucial in mitigating the impact of deepfake content by preventing its dissemination.

Public education campaigns raise awareness about the consequences of creating and sharing deepfake content, aiming to discourage its malicious use such as blackmail or bullying.

Coordinated efforts by policymakers, researchers, technology platforms, and media organisations are vital in tackling deepfake threats and safeguarding a more secure digital environment.

A zero-trust mindset also fosters a proactive defence against deepfakes, promoting healthy scepticism and constant verification of digital content.

Deepfakes pose significant harm to us.

Deepfakes pose a significant threat to society, political systems, and businesses.

Deepfakes generate controversial content such as fake news and pornographic material, threaten national security by spreading propaganda and interfering in elections, and heighten cybersecurity concerns for individuals and organisations.

Deepfakes can also create fraudulent identities in real-time. For example, a worker was deceived into transferring US$25 million to fraudsters who used deepfake technology to impersonate the company's chief financial officer during a video call.

Deepfakes have eroded trust in digital information.

Deepfakes could have far-reaching societal effects beyond mere deception. Deepfakes make it challenging for the public to distinguish between genuine and fabricated news and visual evidence.

These technologies enable the manipulation of genuine audio-visual content, sparking debates about the believability of information disseminated through news and media.

Deepfakes are so realistic that they deceive even those adept at spotting fake content. Consequently, people may lose faith in visual material, leading to the notion that "seeing is no longer believing."

Increased uncertainty about media content may deepen scepticism towards public discourse, journalism, and politics.

So, does this technology have more promising opportunities or pose major risks to society?

AI technology is beneficial to us.

There are AI tools available to detect deepfakes.

AI acts as both a sword and shield in combating deepfake content. AI and defensive tools enable humans to distinguish between authentic content and deepfakes.

For example, Singapore cyber-security firm Ensign InfoSecurity developed Aletheia, a tool that scans for signs of deepfake videos and audio on a user's screen, such as awkward facial and body movements in videos. A similar tool, Einstein.AI, was also developed by Singapore's ST Engineering.

Other tools can also spot AI-generated text, synthesised or altered images, and deepfake content. Integrating multiple detection and analytic tools enhances the identification and elimination of deepfakes.

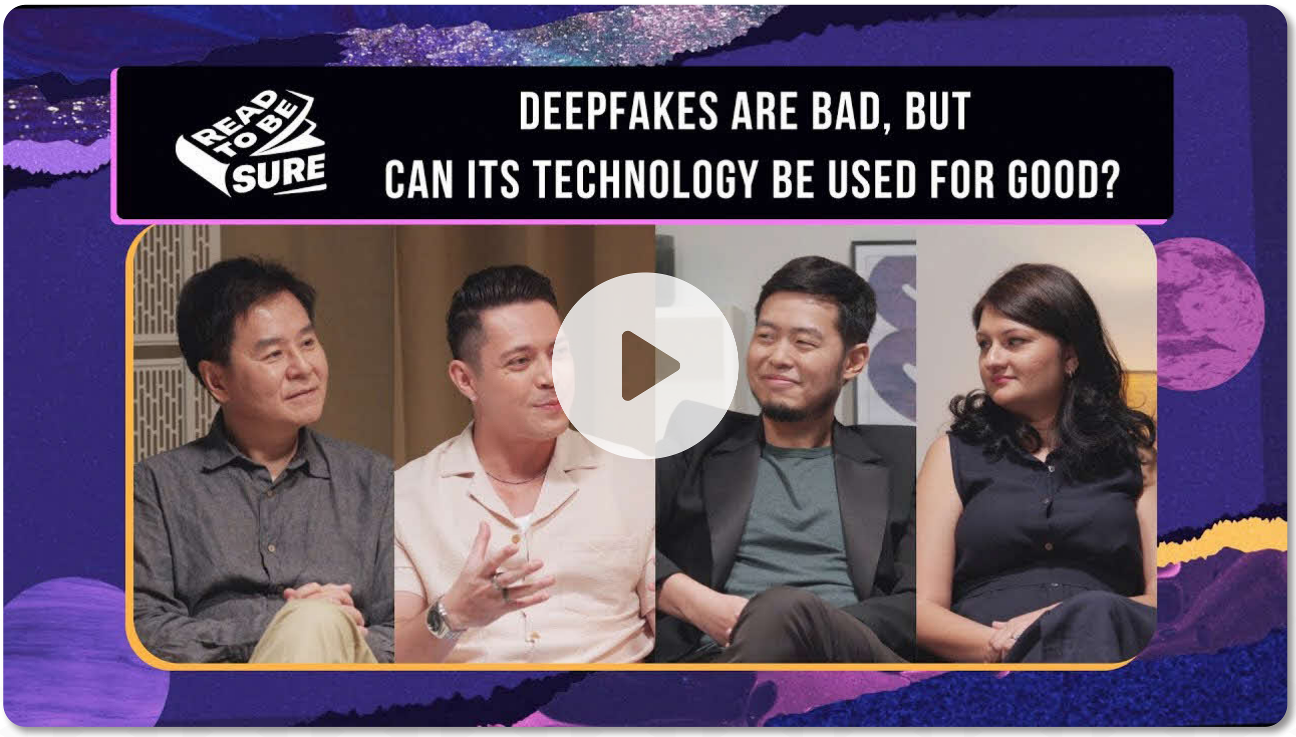
Deepfakes pose significant harm to us.

Most people cannot detect deepfakes reliably.

Humans perform poorly at detecting deepfakes, even when given training on how to spot them. Research has shown low detection rates for both image and audio deepfakes, with accuracy levels of 62% and 73%, respectively, according to articles from the *Journal of Cybersecurity* and *PLOS One.*

Meanwhile, people tend to be overconfident in their ability to identify deepfakes. According to a survey done by Verian, 61% of Singaporeans are confident in their ability to spot deepfakes. This is despite more than three-quarters (77%) expressing concern that deepfakes will be used in scams.

Let's hear our guest speakers weigh in on this issue:



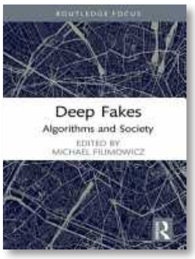
Recommended Resources

Explore the resources below to learn more about Deepfakes.

Video

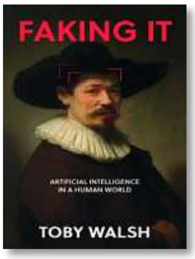
Defense Advanced Research Projects Agency. (2024, March 27). Real or not: Defending authenticity in a digital world. Retrieved 2024, July 12.

NLB eBooks



Deep Fakes
Filimowicz, M. *Deep Fakes*. London: Taylor & Francis, 2022.

Retrieved from OverDrive. (myLibrary ID is required to access the eBook).



Faking It: Artificial Intelligence in a Human World
Walsh, T. *Faking It: Artificial Intelligence in a Human World*. Flint, 2023.

Retrieved from OverDrive. (myLibrary ID is required to access the eBook).

How Algorithms Create and Prevent Fake News
Giansiracusa, N. *Algorithms Create and Prevent Fake News*. Berkeley: Apress, 2021.

Retrieved from OverDrive. (myLibrary ID is required to access the eBook).

Website

Li, C., and A. Callegari. "Stopping AI disinformation: Protecting truth in the digital world". *World Economic Forum*, 14 June, 2024.

Podcast

"The Murky World of Deepfakes". Kathleen McInnis and Di Cooke — *Smart Women, Smart Power*, 15 May, 2024. Podcast, 24:55.

Read to be SURE (RTBS) is part of NLB's Libraries and Archives Blueprint 2025 (LAB25). RTBS aims to educate and inform the public on trending topics through a combination of curated content and accompanying programmes. By exploring multiple perspectives on these topics, readers and participants will benefit from a balanced and credible knowledge base.