THE STRAITS TIMES

Public urged to be on guard as deepfake content will grow more sophisticated: Experts

The spate of deepfake videos underscores the rising threat of AI-generated misinformation. PHOTO: ST FILE

Osmond Chia

UPDATED 9 HOURS AGO

SINGAPORE – To spot a deepfake video, look out for distorted images, speech that does not match the movement of the speaker's lips, and claims that are too good to be true.

Tech experts offered these tips as they called for more education to detect fake videos made using artificial intelligence (AI) after several deepfakes surfaced in December that involved public figures like Prime Minister Lee Hsien Loong.

PM Lee on Dec 29 warned the public in a Facebook post not to respond to scam videos on investments or giveaways after a deepfake video of him purportedly promoting an investment surfaced. It is not known who was involved in the misinformation campaign.

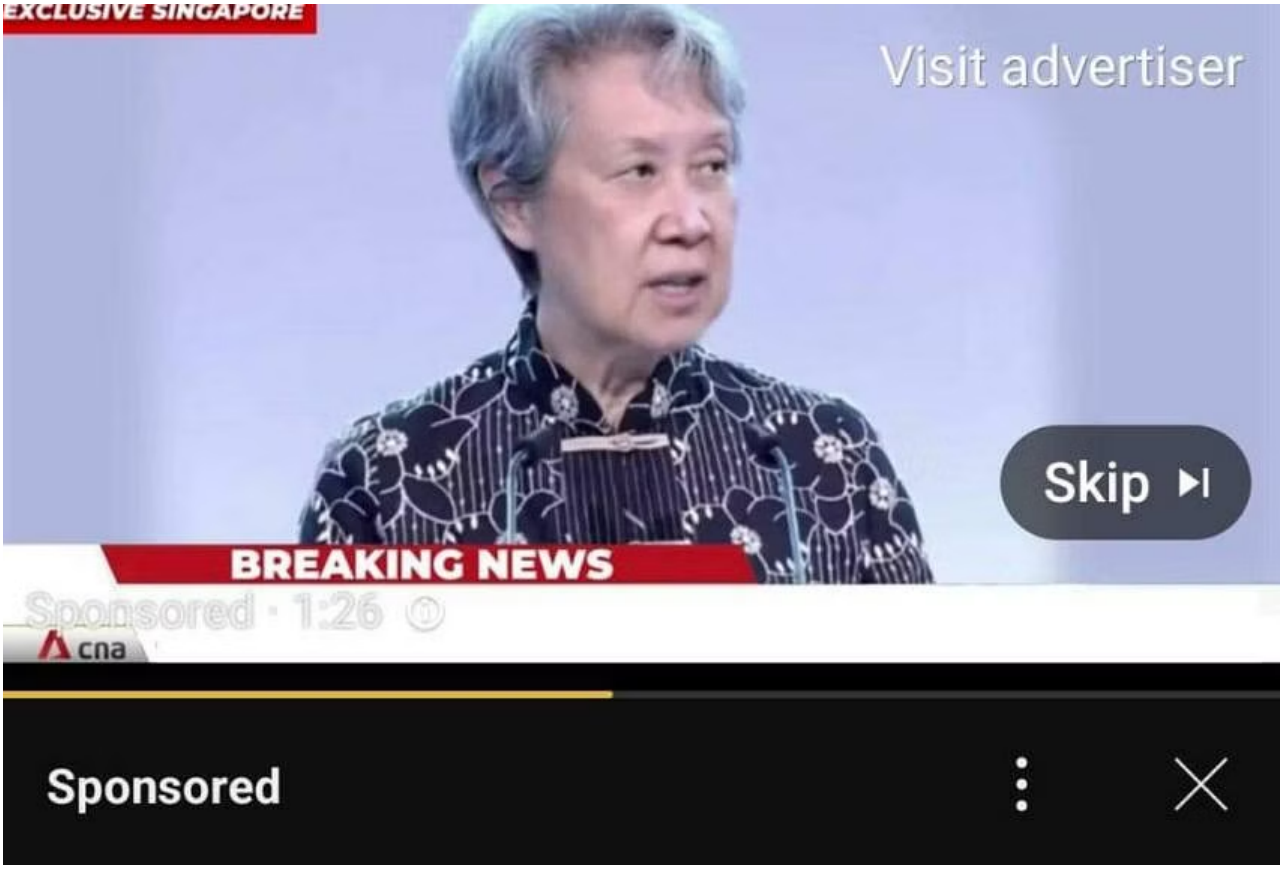
Deputy Prime Minister Lawrence Wong, whose appearance has also been used to promote investment scams, alerted the public on Dec 11 of deepfake posts spreading misinformation that the authorities were planning a circuit breaker amid a spike in Covid-19 cases.

The likeness of Ms Ho Ching – PM Lee's wife and former Temasek chief executive – was also manipulated by fraudsters in another investment video that surfaced in December.

The video circulated as an advertisement on YouTube which plays before users can view the videos they have selected. It is not known who uploaded the false ad.

A spokesman for Google, which runs YouTube, said it has taken down the flagged ad and terminated the associated channel.

Google did not state whether the ad was vetted before it was approved, but said that YouTube will soon require content creators to disclose when they have created realistically altered or synthetic content using AI or other tools. This will be made clear to viewers in the video's description panel, and labelled on the video itself for sensitive content, such as those related to elections, conflicts or public health crises.



A deepfake video of Madam Ho Ching appearing to promote an investment scam had circulated as an ad on YouTube. The flagged ad has since been taken down. PHOTO: ST READER

Deepfakes can be generated if there is a library of footage that enables the AI to capture the facial expressions and verbal cues of an individual. Online AI video generators available to the public can churn out convincing deepfakes, putting anyone's face on a voice in any language, such as pop star Taylor Swift speaking fluent Mandarin.

The technology has been used in films, for instance, to bring back the likeness of dead actors, such as Carrie Fisher as Princess Leia in the newer Star Wars films.

But recent high-profile incidents have shown the dark side of AI-generated images.

The technology caused public outcry in Spain in September after fake images were created of naked underage girls. AI was also used in conflicts, for instance, to generate fake images of injured infants in the Israel-Hamas war, and in 2022 of Ukrainian President Volodymyr Zelensky instructing his soldiers to stand down.

The authorities globally are playing catch-up. In December, the European Union passed its AI Act, which will soon require those who create AI-made videos to watermark their content.

Singapore, too, has introduced legal levers to combat the spread of misinformation online, including the nation's fake news law, the Protection from Online Falsehoods and Manipulation Act, which requires offenders to insert a correction notice against the original post.

The Online Criminal Harms Act, which comes into force in 2024, marks the next milestone in regulating the technology. It will grant the authorities powers to order individuals, firms and Internet service providers to remove or block access to criminal content, including the use of deepfakes for malicious campaigns.

Deepfake videos will become harder to spot with the advancement of technology, Associate Professor Lu Shijian from the Nanyang Technological University School of Computer Science and Engineering told The Straits Times.

Prof Lu's team is similarly developing an AI program that can generate facial animations with just a person's photo and an audio recording of their voice for use in metaverse-related apps and virtual assistants. All content created under the research program, which is called Diverse yet Realistic Facial Animations (Dirfa), is watermarked to make clear that it is not real.

The public needs to be taught how to spot and promptly report such misinformation, said Prof Lu, adding that the authorities and content platforms must do more to penalise those who spread and create malicious deepfakes.

MORE ON THIS TOPIC

PM Lee warns against responding to deepfake videos of him promoting investment scams

Deepfake video of DPM Lawrence Wong promoting investment scam circulating on social media

The spate of deepfake videos underscores the rising threat of AI-generated misinformation, which, at the Singapore Conference on AI for global experts in early December, was named among the dozen most pressing concerns posed by AI.

The experts wrote: "Mis/disinformation is an existing problem, but with the development of AI, we will see an increase in volume and sophistication... We are nearing a point where we lack the ability to discern if the source of information is human or bot, and distinguish between true and fake content."

There are ways to spot AI-made misinformation for now, such as by looking out for distorted movements or mismatches between the audio and the movement of the speakers' lips, said Dr Dimitry Fisher, senior vice-president of data science at Aicadium.

The software company is among at least 60 global industry players that have partnered with the authorities under the AI Verify Foundation to collaborate on AI standards.

"If it sounds too good or too bad to be true, the content you are looking at probably is either outright fake or heavily edited," said Dr Fisher.

"The best bet is to use common sense, and, when in doubt, review the video closely and do additional research."