

Fakes are good business, don't expect AI to stop generating them

By Lim Sun Sun

The Straits Times, Page 3, Section: OPINION | B
Friday 30 January 2026
1266 words, 1089cm² in size
386,100 circulation

Fakes are good business, don't expect AI to stop generating them

Verifying misinformation will only become harder, but consumers can't afford to get complacent.



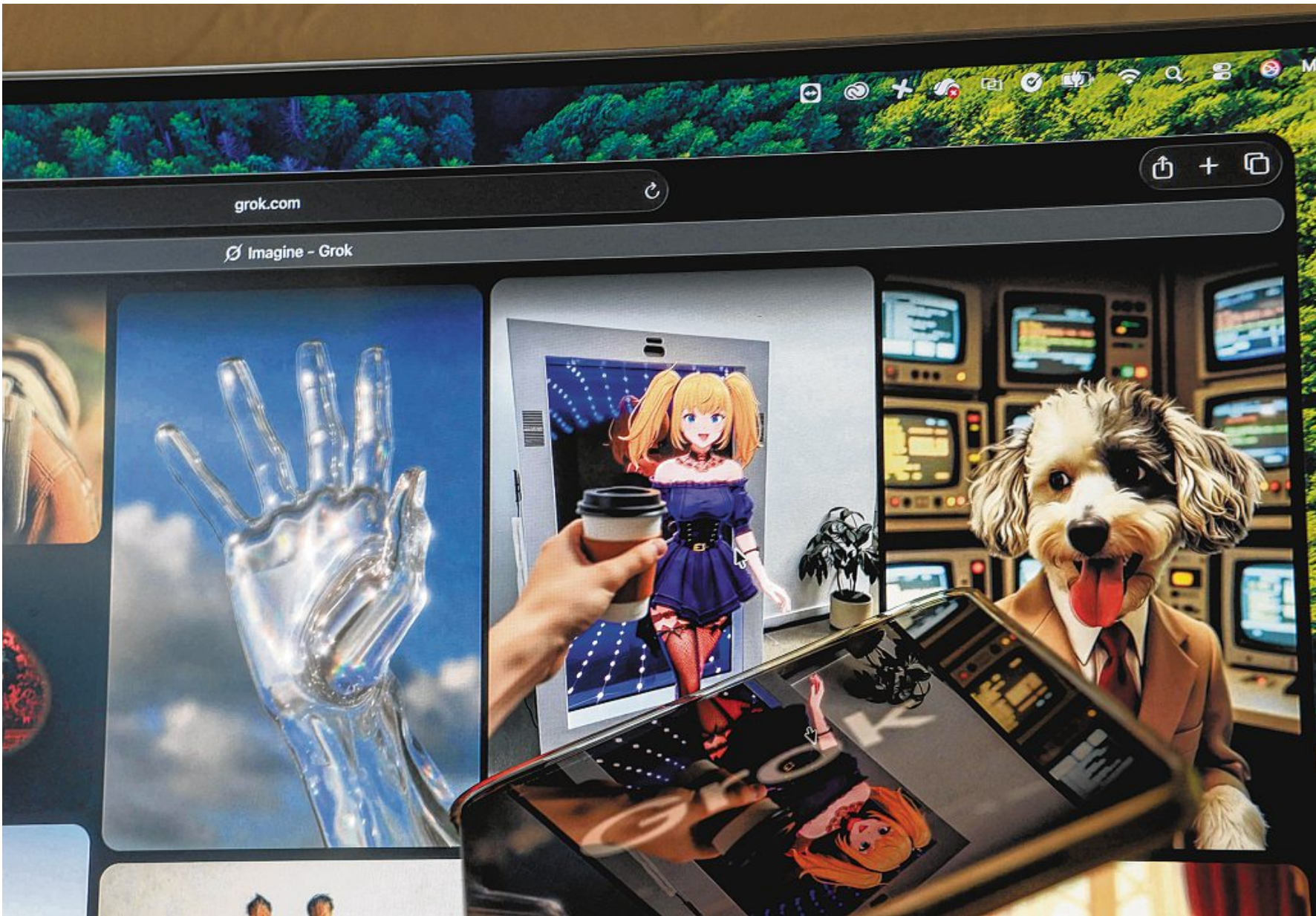
Lim Sun Sun

Hands up if you fell for the AI-generated video of the “emotional support kangaroo” a woman was trying to bring on board her flight, or if an image shared on your WhatsApp chat triggered heated disputes over whether it was AI slop. And if you now find yourself routinely verifying stories that sound so outrageous, an AI tool could have hallucinated them.

Such is our lived reality in a digitally connected world where the capabilities of generative AI have been unleashed without due safeguards. Again and again, we consumers must personally undertake due diligence on content we see or receive, even as we grapple with information overload. But is placing this burden on consumers fair?

In the wake of the new year, we are being deluged with predictions about technological innovations that will shape our daily lives in unprecedented ways for better or for worse. Malware that can autonomously evade detection after infiltrating a device or network, humanoid robots that take out the trash and walk the dog, digital interfaces that detect our nerve signals and convert them into actions – such predictions typically cover frontier technologies that are making their way from lab to living room.

And yet disappointingly few predictions about technological breakthroughs offer solutions to the growing bane of hyper realistic misinformation. Instead, we hear projections that AI-generated content will only improve in quality, shedding current tell-tale signs of people or scenes bearing a veneer of “airbrushed” perfection. Discouragingly therefore, we can expect that misinformation and



The Grok Imagine website is displayed on an iPad and a computer screen. The launch of X's Grok image-generation tool triggered a viral trend of users creating sexually explicit and violent images mostly featuring women. In the absence of reliable misinformation safeguards, the writer suggests that consumers sharpen their discernment, and push tech companies toward more responsible design to restore a healthier information environment. PHOTO: AFP

disinformation wrought by powerful and accessible AI tools will only appear more lifelike, further stacking the deck against feeble consumers.

As it stands, there is a deplorable lag between advances in producing deceptively convincing AI-generated content and robust solutions for evaluating their veracity. Creating a fake video or image using AI image generator tools such as DALL-E or Nano Banana is child's play, but checking the authenticity of images or videos is significantly more laborious.

One has to manually inspect the fine details of the image or video for pixelation, blurred backgrounds, inconsistent shadows, gibberish text or watermarks. Alternatively, one can perform a reverse image search with tools such as Google Lens or TinEye to establish if an image has been manipulated, repurposed from a different context or mislabelled. Checking reputable

news sources for mentions or reproductions of the image or video or reading audience comments on social media posts sharing the content are other means of validation.

Considering the volume of content we consume daily, it is clearly untenable for any consumer to be this vigilant with every image or video they see. And yet, despite the blistering pace of AI developments witnessed in the past three years, we still don't have a reliable falsehood detection tool that works autonomously on our devices and can seamlessly flag and block fake content across all the platforms.

THE BUSINESS OF FAKES

Why are we so far behind the curve? From a technical standpoint, detecting misinformation is a relentless “whack a mole” game where bad actors keep popping up. Although

AI can help identify patterns, it struggles with the nuances of human language, such as sarcasm, satire, or facts taken out of context. Developing tools that can analyse text, images, and video simultaneously and accurately in real-time is a highly complex engineering challenge requiring massive amounts of high-quality data that requires generous resourcing.

Which brings us to the commercial considerations. Given the current business models of online content, misinformation is more profitable than the truth by a mile. Sensationalist “clickbait” drives high engagement, which generates advertising revenue for platforms, creating a financial conflict of interest when it comes to aggressively removing content.

Additionally, it is incredibly cheap and fast to generate misinformation and deepfakes using AI, but extremely costly and time-consuming to build the sophisticated systems needed to

catch them. Companies are also wary of the legal and reputational risks; if a detection tool accidentally flags a real news story as “fake,” it can lead to charges of censorship or even crippling lawsuits.

The recent uproar over X's Grok image-generation tool offers a stark example of how commercial incentives can undermine content policing. The launch of the tool triggered a viral trend of users creating sexually explicit and violent images mostly featuring women. Alarming, the Center for Countering Digital Hate estimates that Grok AI generated about three million sexualised images in 11 days.

In response to the backlash, X moved these image-generation features behind a paywall that critics say turns abuse into a premium product. This disturbing development also exposes a familiar pattern of reactive design, where minimal safeguards allowed harmful content to spread at scale

before regulators could step in, suggesting a rush to market with safety thrown by the wayside.

By framing the issue as one of “free speech” and branding the tool as deliberately “rebellious,” X also sought to shift responsibility onto users while sidestepping the costly work of robust moderation. The Grok episode reveals what is by now a well-worn trope: without sustained external regulatory pressure, tech platforms are structurally incentivised to prioritise novelty and subscription revenue over proactive content policing.

SO WHAT CAN WE DO?

Which leads us to return to consumers and our individual roles in fortifying the information environment. Without the assurance of failsafe misinformation detection tools at our fingertips, we need to sharpen our skills of discernment. Be selective about our news and information sources, be judicious about what we share and be conscious of our own cognitive biases that make us more vulnerable to falsehoods.

In today's dense and chaotic media environment, we should also normalise calling out or questioning content that appears to be false and AI-generated. And when it comes to generating media content of your own, don't indiscriminately use AI for everything but continue to use your distinctively human skills to sharpen them.

On their part, tech companies must stop bludgeoning us with more AI tools and services that we do not actually need. Instead, they should look for legitimate solutions to address problems such as the scourge of AI-generated misinformation.

As society buckles under the growing weight of falsehoods and deepfakes, a safe and secure information platform would present a viable value proposition that could even reap commercial gain. Policy innovations may be needed to disrupt and reinvent current business models, although the current geopolitical climate does not bode well for prosocial approaches.

Consumers shouldn't be left to flounder in the choppy waters of AI-generated misinformation, but neither can we afford to be complacent. By demanding more responsible design from tech companies and being more discerning ourselves, we can at least tilt the scales back toward a healthier information environment. If AI is here to stay, then we too must insist that it serves the public good rather than eroding it.

• Lim Sun Sun is vice-president, partnerships and engagement at Singapore Management University and Lee Kong Chian professor of communication and technology at its College of Integrative Studies. Her latest book is Humanising Technology: Reflections on Design, Ethics and Inclusion.