

Harmless fun? People are using Sora to create problematic videos

Hyper-realistic videos that mock certain groups or reinforce harmful stereotypes can shape attitudes over the long term.

**Elvin Xing
and Sophy Tio**

A cat ice-skating. Historical figures playing basketball. Cartoon characters hanging out with famous people.

Generative video tools like OpenAI's Sora 2 mark the next frontier in artificial intelligence (AI) and are pushing the boundaries of realism. This ability to reimagine reality and to create videos can be empowering and engaging.

But alongside these playful clips, more troubling genres are emerging: hyper-realistic videos that mock certain groups, distort real events, or reinforce harmful stereotypes.

HARMLESS FUN OR NORMALISING CONTEMPT?

Consider the wave of AI videos depicting overweight individuals stumbling, eating excessively or engaging in derisive behaviour earning them mockery. These arise from problematic prompts, yet are framed and shared as light-hearted entertainment.

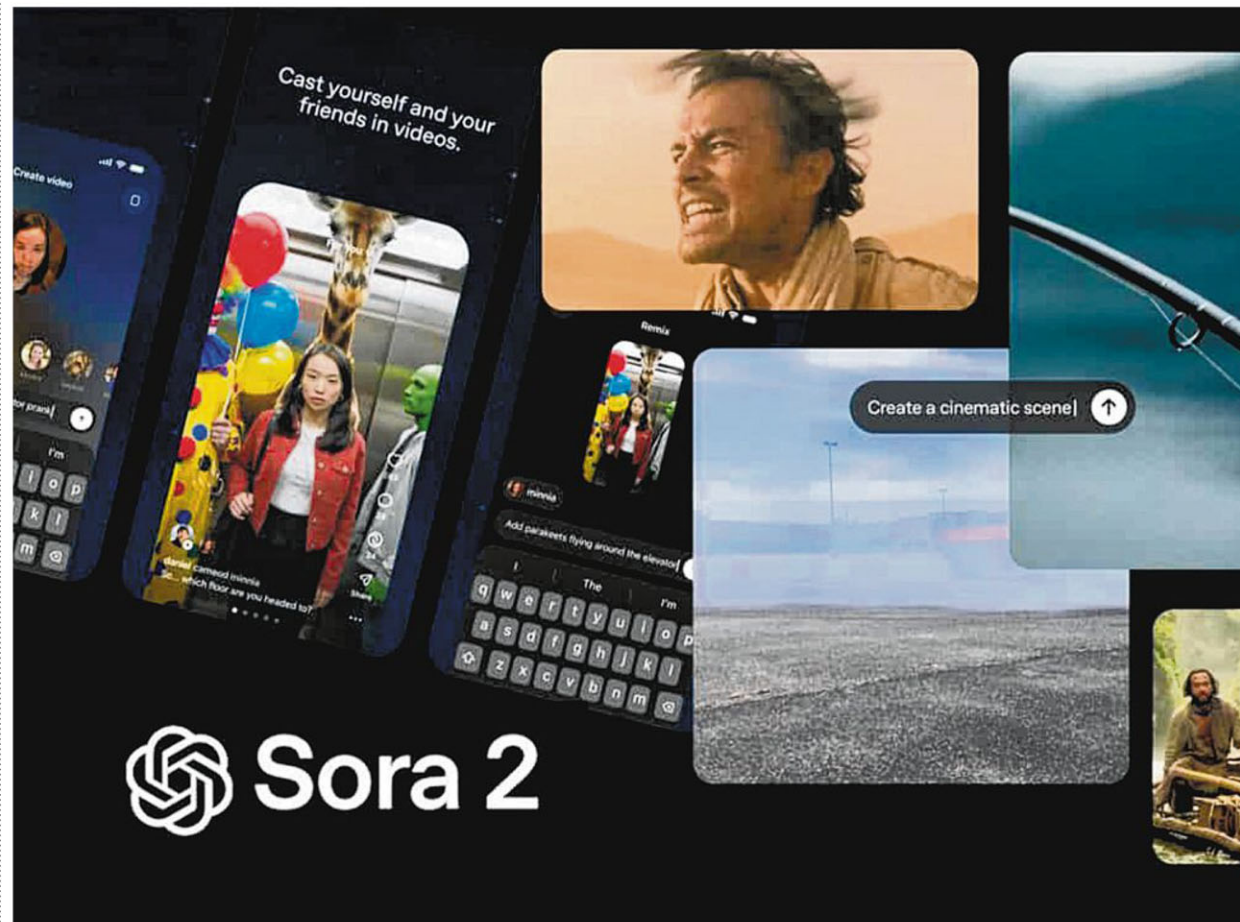
But each view, like or share subtly reinforces the idea that certain bodies, identities or cultural expressions deserve to be ridiculed, broadcasting microaggressions at scale.

As research from MIT shows, sensational content travels faster and further than factual content, a dynamic that generative videos on algorithmic social media accelerates. A single offensive joke told in a room is one thing; an AI-generated video replicated millions of times reshapes perceptions, conditioning audiences to accept contempt as normal or even humorous.

The line between humour and problematic stereotyping is not always clear. But the comedy industry has a term for jokes that denigrate people: "punching down" or mocking a group more vulnerable than oneself. It is generally seen as bullying, not humour.

In the US, where Sora 2 was first launched, a slew of videos portraying black women as angry, low-income and dissatisfied with government benefits have circulated widely.

Even when such videos are



Generative video tools like OpenAI's Sora 2 are pushing the boundaries of reality. Over time, the relentless exposure to content that reinforces harmful stereotypes can erode empathy, shrink spaces for dialogue and deepen divides. What starts as a meme can become a method of exclusion and marginalisation. PHOTO: OPENAI

watermarked as AI or fleeting in duration, they quickly spread disinformation and reinforce damaging stereotypes of black communities as emotional and poor. Over time, the continuous stream of content shapes public attitudes, spreads ignorance and even evolves into hate.

This is the machinery of "othering" at scale, and it seeps into offline life.

It often starts off with jokes dismissed as "just kidding", memes passed off as mindless humour, or stereotypes excused as exaggeration. This is not an abstract issue in Singapore.

The IPS Working Paper on Prejudice, Attitudes and Critical Perspectives on Race in Singapore published in July found that over one-third of Malay and Indian respondents reported personally experiencing jokes based on race or religion far more often than Chinese respondents.

Racism is not only expressed in overt acts of discrimination, but also embedded in the subtlety of backhanded compliments such as "you're smart for your kind" and "benign" jokes that serve to typecast entire groups with a

single characteristic.

Such seemingly minor acts can create a pervasive environment that disproportionately affects minority groups, gradually eroding the trust that underpins Singapore's multicultural compact, where our social fabric is so deeply interwoven with a diversity of racial and religious identities.

Offline or online, these jokes fuel negative perceptions of the "other" and influence the way we interact with other groups in public, and at our workplaces.

In research looking at memes and political discourse, one study in Germany shows far-right actors using humour and ideology embedded in social media content. These work like Trojan Horses to co-opt broader audiences into the "in-group" and sow discord with the "out-group" via indirect hate speech that deprecates them.

Over time, the relentless exposure to such content can erode empathy, shrink spaces for dialogue and deepen social divides. What starts as a meme can evolve into a method of exclusion and marginalisation.

DRAWING LINES

We must therefore draw a crucial distinction between humour that uplifts and humour that degrades and guard against divisive content that denigrates people on the basis of race, religion, ethnicity or other markers of identity.

The solution is not as simple as training AI to correct systemic bias. In 2024, Google had to issue an apology after its Gemini AI tool wrongly depicted the race of white historical figures as people of colour. Such overcorrections risk introducing new biases and distorting facts even as AI attempts to reflect racial diversity.

What we need are systems that understand social context, that distinguish the difference between empowering representation and harmful distortion.

Just as Singapore draws legal and moral boundaries around hate speech offline, we must establish similar guardrails online. Laws such as the Maintenance of Religious Harmony Act, the recent Maintenance of Racial Harmony Act, and the Online Safety (Relief and Accountability)

Act provide mechanisms to mitigate the generation and spread of offensive material online, but subtler stereotyping, memes or AI-generated caricatures still fall into the grey area of harmful content.

Meanwhile, the Infocomm Media Development Authority's (IMDA) Model AI Governance Framework for Generative AI offers guidance for responsible and ethical AI development and deployment, hopefully helping to reduce the generation and amplification of divisive content.

Yet legal framework and tech safeguards are insufficient. Digital literacy and proactive community interventions remain essential.

BUILDING LITERACY

Addressing the risks requires a strategy grounded in building community resilience against online harms.

First, digital literacy must be treated as a core competency in Singapore. As AI-generated content becomes increasingly indistinguishable from reality, people must be equipped to critically interrogate what they

encounter online.

At an institutional level, Character and Citizenship Education in schools, digital literacy programmes for seniors such as the IMDA Seniors Go Digital programme and the IMDA's Digital Skills for Life framework should teach people to pause and ask: Who created this? Is this historically or socially accurate? Can it be verified? Could it reinforce stereotypes?

These initiatives must also incorporate sustained awareness-building to help users identify bias in digital content and recognise their own unconscious biases. Such skills help not only consumers but also creators to assess the potential harm their content might cause, understand how unchecked narratives can contribute to division and hold them accountable for their content.

Second, research on contact hypothesis shows that positive face-to-face intergroup interactions reduce prejudice and negative attitudes towards others. Studies in the UK show this principle extends to online and indirect contact as well.

Social environments where intergroup interactions occur, such as schools, workplaces, community centres and neighbourhoods, can be intentionally shaped to include positive digital interactions. For example, team-bonding activities can involve co-creating online content that highlights diversity and collaboration.

Such environments can also serve as training grounds for constructive online engagement. Workshops like the IPS Programme on Race, Religion and Intergroup Cohesion emphasise intervention skills for responding to inflammatory content targeting individuals or groups. These training sessions enable netizens to de-escalate online conflicts, challenge prejudice constructively and mitigate the impact of harmful online stereotypes.

Finally, we should use technology intentionally to strengthen social cohesion. In a diverse society like Singapore, where trust across racial and religious lines is carefully built, AI must be used to highlight shared values, uplift underrepresented voices, and foster empathy.

This includes incentivising educators, creators and platforms to prioritise inclusive narratives that promote positive cross-cultural interactions and rewarding systems that represent diverse groups fairly and respectfully. The aim is not to limit the use of AI, but to channel it towards supporting a cohesive society.

Generative AI is a mirror, reflecting our best and worst impulses. We should not douse the joy in using these tools, but must keep in mind the values we hold dear as a society – cohesion, multiculturalism and respect for all.

• Elvin Xing is research fellow at the Institute of Policy Studies Social Lab, National University of Singapore. Sophy Tio is research associate at the same institute, and is a lead facilitator and coordinator of the IPS Programme on Race, Religion and Intergroup Cohesion.