



The Trouble When Fact-Checking is In English and Social Media Isn't

Benjamin Ang



RSIS Commentary is a platform to provide timely and, where appropriate, policy-relevant commentary and analysis of topical and contemporary issues. The authors' views are their own and do not represent the official position of the S. Rajaratnam School of International Studies (RSIS), NTU. These commentaries may be reproduced with prior permission from RSIS and due credit to the author(s) and RSIS. Please email to Editor RSIS Commentary at RSISPublications@ntu.edu.sg.

The Trouble When Fact-Checking is In English and Social Media Isn't

By Benjamin Ang

SYNOPSIS

Sensational AI-generated Chinese-language videos are disseminating misleading claims concerning Singapore, capitalising on less stringent moderation in non-English platforms. This linguistic disparity risks fragmenting public understanding. Strengthening multilingual monitoring, using regionally trained AI instruments, and empowering trusted community representatives are essential measures to combat misinformation and safeguard social cohesion.

COMMENTARY

If you usually read and watch your news and entertainment in English, then there's a hidden battle for truth going on in the digital battlespace that you may not have noticed. In recent months, there has been a wave of sensational Chinese-language videos on social media platforms, claiming that Singapore's political leadership is in a state of "turmoil" or "internal strife". With alarming titles such as "Singapore is starting to bleed" and "The chaos in Singapore", these videos are not just clickbait, but can also erode public trust in public institutions and the economy, thereby threatening national security.

Researchers have identified a low-cost assembly line behind this surge, with threat actors leveraging generative AI tools such as DeepSeek or Ernie to automate the entire production process – from scriptwriting and voice-overs to automated video editing with captions – for as little as US\$1 (S\$1.30) to US\$2 for each 20-minute video. The technology has enabled channels on YouTube and TikTok to churn out hundreds of videos in the past few months, splashing sensational and misleading stories about regional politics.

Our team examined these channels and found them to be a motley collection that shared certain common threads. Some recycled stock videos or old TV footage with rapid-fire voice-overs and captions, while others have a single influencer speaking to the camera, but they often share the same script verbatim. Their content is not explicitly illegal, and therefore not subject to regulation, but it is sensational and misleading. Creators may be motivated by political agendas, advertising revenue, or the desire to build and exploit a loyal audience for future financial scams.

The Linguistic Blind Spot

Singapore faces a linguistic vulnerability that is also seen elsewhere: while English-language content is robustly moderated, vernacular content remains under-monitored. Research from the Harvard Kennedy School Misinformation Review highlights that most monitoring and debunking of disinformation is done in the languages of high-income Western countries. This holds true in Singapore for English, too, and leaves our other official languages (Mandarin, Malay, and Tamil, also known as “mother tongue languages”) relatively under-reported and under-served.

In the context of national security, this linguistic gap is a structural weakness. As the majority of experts monitor the information environment primarily in English, we miss the early warning signs of hostile information campaigns circulating in other linguistic communities.

Disinformation thrives in these spaces because the automated filters of global tech platforms like Meta are notoriously less effective at catching nuances, slang and cultural context in non-English content. Reports have shown that platforms often fail to stop dangerous disinformation, even in the world’s most spoken languages, simply because they lack the localised linguistic expertise to distinguish between legitimate political discourse and coordinated inauthenticity.

This is not just an issue of language, but also of social divisions in our multicultural nation. According to the 2020 Population Census, while nearly half of residents speak English the most at home, there are differences in usage of mother tongue languages, which are linked to age, education levels, socio-economic status, and immigration status.

Studies by the Institute of Policy Studies have shown that language proficiency shapes identity and the way residents consume news. If residents who rely on Mandarin, Malay, or Tamil sources for their primary information are greatly under-served by fact-checking resources and research, then there is a risk that different segments of our population may believe vastly different versions of events based on the linguistic cyberspace they inhabit. This, in turn, undermines the national consensus required for social stability.

The Disconnect in Debunking

Even when a malicious post is successfully detected, the response must overcome the language gap. For example, if a fake Chinese-language video goes viral online, the subsequent official debunking cannot be limited to the English news media.

There would be a fundamental psychological disconnect if consumers' primary information source is an exciting, sensational video in Mandarin, but the response is a factual, official English-language press release. The correction may never reach them, or threat actors can frame it as "government suppression" of a "truth" that only the Mother Tongue Language audience is "brave enough" to hear. This can play into the hands of attackers who use online falsehoods to build conspiracy narratives around the "English-speaking elites" establishment versus the "neglected" vernacular speaker.

That is why, to strengthen Singapore's digital defence, we must move beyond reactive, English-centric rebuttals. We need a strategic approach that centres on community and language.

First, we must expand multilingual monitoring and fact-checking capabilities. This requires building up experts with cultural intelligence to understand the nuances of how narratives apply to different groups. We should also be using AI tools to fight AI slop and deploy multilingual large language models developed in Asia to detect and flag sensational rhetoric in non-English content as it appears, rather than waiting for it to go viral.

Second, there is an important role for key opinion leaders and community influencers who are proficient in mother tongue languages and trusted within their respective circles. These trusted speakers can debunk misinformation in the same language and on the same platforms (like WhatsApp, WeChat, or Telegram) where the misinformation originated. They can also help their communities understand the motives and tactics of such videos before they appear, so that everyone can be a critical consumer of information, regardless of language.

Singapore's resilience in the digital age depends on our ability to protect a common understanding of facts that is not limited by race, language, education or social status. Our strongest defence against evolving threats is a shared national identity, not silos of separate linguistic realities. By strengthening our multilingual defences, we deny hostile actors the chance to exploit our diversity as a weakness.

Benjamin Ang is Head of the Centre of Excellence for National Security and Future Issues and Technology at the S. Rajaratnam School of International Studies (RSIS) at Nanyang Technological University (NTU), Singapore. This commentary was originally published in The Straits Times on 26 February 2026. It is republished here with permission.

Please share this publication with your friends. They can subscribe to RSIS publications by scanning the QR Code below.

