

October 14, 2026 Responsibility & Safety

Google DeepMind and Singapore Government collaborate to build fine-tuned AI model for scam detection

JJ Geewax and Miguel de Andrés

GovTech and Google DeepMind build on-device AI that warns Android users of scam chats

Description

A new artificial intelligence model built by GovTech Singapore and Google DeepMind will spot scam chats on Android phones as they happen and warn users before they hand over money or passwords. The model runs on the phone itself, so messages never leave the device.

October 14, 2026 Responsibility & Safety

Google DeepMind and Singapore Government collaborate to build fine-tuned AI model for scam detection

JJ Geewax and Miguel de Andrés

GovTech Singapore and Google DeepMind fine-tuned the scam detection model on local scam conversations. Screenshot: Google DeepMind, taken 18 October 2026

The model is a version of Google's lightweight Gemma model, fine-tuned on local scam conversations. These range from impersonations of public institutions, banks and local figures to trending regional schemes, so the model can tell ordinary chatter from genuine manipulation.

How the warning works

The model runs entirely on the phone rather than scanning messages on a cloud server. That matters because many messaging apps use end-to-end encryption, which keeps private conversations out of reach of server-side checks.

Detection runs in two stages to save battery. A small, low-power classifier screens incoming messages first. If it spots a possible threat, the fine-tuned Gemma model reads a larger slice of the conversation to confirm the scam or discard a false alarm. A confirmed threat triggers a real-time warning on screen, telling the user to pause before sharing money or sensitive details.

Privacy and rollout

The model will reach users inside partner apps, in a phased rollout developed with the Singapore Government. It can only read on-device conversations with the user's consent, under that app's terms of use. Neither the Government nor Google can see those conversations.

Training took place in a GovTech-managed Google Cloud environment, using anonymised data from scam reports and intelligence feeds rather than personal information. The model was stress-tested against synthetic dialogues and local scam corpora.

"Fighting scams shouldn't mean sacrificing personal privacy," said Mark Chen, director of trust and safety at GovTech Singapore.

Why scams are the target

Scams make up more than half of Singapore's reported crimes, with losses of more than S\$4 billion since 2019. Social engineering scams, in which victims are drawn into convincing conversations before being asked for money, account for up to 80 per cent of losses.

A single scam can run from social media to a messaging app, a phone call and a bank transfer. "If our devices can recognise signs of manipulation early, we can help potential victims avoid financial or credential loss, while keeping their personal messages on their devices," Minister for Digital Development and Information Josephine Teo said at Singapore International Cyber Week on 14 October.

The Government already uses AI against scams elsewhere in that chain. The Scam Analytics and Tactical Intervention System, built by GovTech and the Singapore Police Force, scans more than 600,000 websites a day and disrupts 20,000 malicious sites a month.

Date Created

18/10/2026

Author

vanessakoh