RSIS COMMENTARY

28 May 2026

Why Agentic AI Social Networks are a Southeast Asian Concern

Karryl Kim Sagun Trajano and Ysa Marie Cayabyab

RSiS S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES
Nanyang Technological University, Singapore

RSiS 30

RSIS S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES
Nanyang Technological University, Singapore

RSIS COMMENTARY

www.rsis.edu.sg No. 115 – 28 May 2026

Why Agentic AI Social Networks are a Southeast Asian Concern

By Karryl Kim Sagun Trajano and Ysa Marie Cayabyab

SYNOPSIS

Meta's acquisition of Moltbook signals a broader transition from AI as a human-assisted tool to systems capable of acting autonomously across digital ecosystems. This commentary examines how AI agent swarms could reshape Southeast Asia's information, economic, and security environments, promising efficiency gains and deeper digital integration, but also creating new vulnerabilities linked to misinformation, cybercrime, market manipulation, and governance gaps.

COMMENTARY

Meta's acquisition of Moltbook, an autonomous social network of AI agents, shows how far AI governance frameworks must evolve. Moltbook represents a novel form of online environment in which AI agents interact and generate content at scale.

For policymakers in Southeast Asia, the development poses a dilemma: technological experiments originating in global tech hubs can rapidly diffuse across regional digital systems, often outpacing existing regulatory oversight.

The emergence of AI agents highlights a transition from AI as a human-assisted tool to AI systems capable of acting autonomously on users' behalf. AI agents can operate continuously and independently, interpreting and responding to content at scale while autonomously completing tasks. In shared environments, agents can detect bugs, optimise workflows, and solve problems faster than humans. With elevated permissions and delegated authority, AI agents are already carrying out actions with

consequences, from sending messages in a user's name to initiating financial transactions.

Unlike human-directed AI, autonomous agents can amplify security risks more quickly and at a greater scale. Swarms of agents can generate disinformation, amplify narratives, and simulate public consensus. Cybercrime can also be automated, with agents probing vulnerabilities, launching phishing campaigns, and coordinating fraud. In the Moltbook data breach, 1.5 million authentication keys and identities were exposed, potentially allowing attackers to manipulate AI agents into extracting or deleting sensitive data.

But existing regulations were designed for human users, not autonomous AI agents. If AI agents cause financial, social, or physical harm, legal liability remains unclear. Without standards for identity, transparency, and accountability, agent-based systems risk becoming opaque and easily exploited.

AI agent swarms could also generate and spread localised narratives in multiple languages, deepening polarisation and undermining trust. Given its linguistic diversity and political sensitivities, ASEAN provides a fertile ground for automated disinformation.

Likewise, the high level of regional digital integration and cross-border e-commerce infrastructure create conditions in which malicious agents could generate economic and cybersecurity risks, such as fraud automation and transaction manipulation.

Unfortunately, governance responses are uneven across the region. While Singapore has advanced AI governance frameworks, many ASEAN states still face capacity gaps in cybersecurity, technical expertise, and monitoring. This creates weak links in regional digital ecosystems that high-risk platforms could exploit.

Beyond cybersecurity and governance challenges, platforms, including Moltbook, also raise societal and political concerns as autonomous agents may fail to internalise norms surrounding racial and religious harmony. Without safeguards, agents could amplify misinformation, bias, or polarising narratives. Their human-like communication makes such content more persuasive and harder to detect, contributing to machine-driven echo chambers, including trivial but harmful fabrications such as invented religions.

These risks expose limits in current governance frameworks, which emphasise transparency, accountability, safety, and human oversight. While Singapore has introduced the world's first agentic AI governance framework, platforms such as Moltbook operate across transnational architectures that diffuse responsibility, making accountability harder to assign.

The Moltbook case is an early example of large-scale agentic systems in the wild – a chance for regulators to identify and address blind spots before such systems become embedded in finance, governance, and trade. It also shows the need to adapt existing ASEAN soft-law instruments on AI, cybercrime, and online harms to treat agentic AI as a shared digital infrastructure that requires coordinated oversight.

Moltbook also shows that agentic AI operates across borders, making regional governance alone insufficient. ASEAN must complement its efforts with global partnerships, including regulatory dialogue with the EU and OECD, and by drawing on frameworks such as the EU AI Act and Digital Services Act for risk and accountability models, and the OECD AI Principles for global norms.

Financial regulators can work with the Bank for International Settlements Innovation Hub to assess systemic risks from agents in payment systems and digital assets. ASEAN could also support global standards for agent identity, authentication, and logging through bodies including the National Institute of Standards and Technology, ISO/IEC 42001:2023, and Singapore's Artificial Intelligence Technical Committee. Cooperation with INTERPOL and EU cyber networks can likewise improve fraud detection, intelligence sharing, and coordinated responses.

Cross-regional research sandboxes could also explore agent safety and human override mechanisms to ensure that agents can be stopped before unintended actions, harmful narratives, or autonomous decisions escalate beyond human control. Singapore's Infocomm Media Development Authority (IMDA) recently warned users against granting the AI tool, OpenClaw, unrestricted access to sensitive data. IMDA also advised organisations to review OpenClaw's deployment in mission-critical environments, including core production and financial systems, highlighting growing security concerns.

This points to the need for co-creation with other regions – North America, Europe, and East Asia – to test least-privilege configurations, human override, and deception scenarios, and positioning Southeast Asia as a contributor to next-generation safety architecture.
The Moltbook case makes the case for rethinking AI governance, shifting from content moderation to system design, given agentic AI relies on permissions, autonomy, and machine-to-machine interaction. Regulation must move upstream through identity standards, access controls, logs, and embedded liability.

The challenge for Southeast Asia is not whether agentic AI will shape its digital future, but whether the region can meaningfully take part in shaping the rules, norms, and safeguards governing it.

Karryl Kim Sagun Trajano is a Research Fellow at Future Issues and Technology (FIT), S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore. Ysa Marie Cayabyab is an Associate Research Fellow at FIT, RSIS. *This commentary was originally published in The Interpreter (Lowy Institute) on 26 May 2026. It is republished with permission.*

S. Rajaratnam School of International Studies, NTU Singapore
Block S4, Level B3, 50 Nanyang Avenue, Singapore 639798

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