Council for Security Cooperation in the Asia Pacific

Regional Security Outlook

2026

The AI Paradox:

Generative AI and Scams in Southeast Asia

Mark Bryan Manantan, Research Fellow, Centre for Global Security, La Trobe University

With the anticipated launch of the ASEAN Digital Economy Framework Agreement during the Philippines' chairmanship in 2026—alongside the implementation of Regional Payment Connectivity to promote effective cross-border digital payments—combatting online scams has become an urgent regional priority for Southeast Asia. The same forces accelerating digital integration and economic opportunity are also generating new and expanding vulnerabilities.

Transnational crime syndicates are adapting quickly to Southeast Asia's technological drive. The advent of generative artificial intelligence (AI) has further accelerated the cyber-scams industry, equipping malicious actors with increasingly novel ways to refine their tactics and scale deception across borders, both offline and online. In response, regional and global discussions have converged on the Global Public-Private Partnership Framework against Fraud as a key mechanism to disrupt and counter the evolving service models of cyber scammers. The central challenge, however, lies in whether these coalitions can respond rapidly and coherently enough before transnational crime syndicates adapt once again.

This article examines the dual role of AI—particularly generative AI—in amplifying and mitigating the rise of cyber scams in Southeast Asia. It outlines key implications and policy recommendations for countering cyber-scam operations, with a focus on strengthening regional information sharing and extracting tangible and measurable contributions from AI companies to prevent the misuse of their models.

Online Scams in Southeast Asia

With rapid digitalization and rising internet penetration, Southeast Asia now sits in the front row of the global online scams industry—both as a primary target and a major operational hub (Sari 2014). Across the region, ASEAN member states have articulated ambitious strategies to attract investments for AI-enabled systems and solutions, semiconductors, data centres, cloud services, and undersea cables. At the same time, Southeast Asia's rapid transition toward mobile-first economies has spurred the widespread adoption of mobile-banking applications and e-commerce platforms (Manantan 2023; Raman et al. 2024). Cyber scammers have exploited such trends.

> “Since the pandemic, Southeast Asia has recorded an annual increase of roughly 50 percent in scam incidents, with total losses approaching US$5 billion”

President Marcos pushes for stronger coordination, implementation of ASEAN plans to mitigate impact of future global shocks, May 9, 2026. Presidential Communications Office, Public domain, via Wikimedia Commons

Since the pandemic, Southeast Asia has recorded an annual increase of roughly 50 percent in scam incidents, with total losses approaching US$5 billion (Raman et al. 2024). According to the United Nations Office on Drugs and Crime (UNODC), organized crime groups operating from Southeast Asia accounted for an estimated US$37 billion in online-scam losses across East and Southeast Asia in 2023 alone (UNODC 2025b). The exponential growth of these losses underscore the adaptability of transnational crime syndicates. Evidence suggests that many Asian criminal networks have pivoted seamlessly from illegal online-casino operations to cyber-enabled fraud, including investment scams, romance scams, phishing, and crypto-based money laundering (UNODC 2025b).

The Mekong subregion—comprising Cambodia, Laos, Myanmar, Thailand, and Vietnam—emerged as a central hub of scam activity in terms of scale, pace, and sophistication between 2023 and 2025. Despite intensified crackdowns led by UNODC and INTERPOL, criminal networks continue to thrive by exploiting the region’s porous borders and uneven law-enforcement capabilities. Scam operations remain highly mobile, relocating operations quickly across jurisdictions—a trend reflected in confirmed spikes in the cross-border movement of labour and illicit activity (UNODC 2025b).

These adaptations extend beyond national borders. Scam syndicates have diversified their digital infrastructure, increasingly leveraging satellite-internet services, such as Starlink, to sustain operations in remote and weakly governed areas (UNODC 2025a). Law-enforcement agencies have also observed the emergence of self-contained digital ecosystems for laundering illicit proceeds, built around online-payment applications, encrypted-messaging platforms, and cryptocurrencies that bypass mainstream financial systems (UNODC 2025a).

Raided gang-run internet 'scam farm' in Bamban, north of Manila, the Philippines, May 31, 2024. Photo by United Nations Office on Drugs and Crime via Wikimedia Commons

Exacerbating these developments is the widespread trafficking of individuals for forced criminality. Scam compounds increasingly rely on trafficked workers of diverse nationalities, moving beyond their earlier dependence on mainland Chinese nationals. Investigative reporting has documented thousands of individuals across Southeast Asia, South Asia, and Africa lured by fraudulent job offers and subsequently trapped in scam-operation compounds (UNODC 2025b).

The economic and social consequences of scams are significant. Online scams have eroded confidence across Southeast Asia, imposing heavy financial and emotional costs on victims. According to the *ASEAN Consumer Scam Report 2025,* more than 67 percent of consumers are "very worried" about online scams while 84 percent fear that the threats are intensifying (GSMA 2025). These developments are already reshaping consumer behaviour, with over 90 percent of respondents adopting more cautious approaches to e-commerce—raising concerns for a region where digital trade remains a key growth engine (GSMA 2025).

Is AI a Friend or Foe?

Despite progress in regional co-operation led by UNODC and INTERPOL, countering online scams has become more daunting with the rapid integration of frontier AI models (like ChatGPT, Gemini, and Claude) into scam-enabling tools that enhance deception and circumvent platform restrictions (Girolamo 2026). UNODC has characterized large language models (LLMs) as a force-multiplier for existing criminal activities, lowering barriers to entry, while amplifying scale, speed, and reach (UNODC 2024b).

Low-cost access to generative AI, coupled with expanding connectivity and fragmented regulatory environments, has enabled scam operators to automate and refine fraudulent practices at unprecedented speed. At the 2026 UN Global Fraud Summit in Vienna, UNODC Acting Executive Director John Brandolino (2026) cautioned that "with today's technology, everyone is a potential target, and no target is out of reach. Digital tools have reshaped the fraud landscape in critical ways, and we need to rapidly step up our efforts and take action to fight back." The discussions below offer brief insights into the growing application of LLMs in the scams ecosystem and how governments are deploying them too, as countermeasures.

Sharpening the Saw: AI-Driven Scam Tactics

Criminal groups increasingly deploy AI to generate fake identities, produce phishing content, and craft highly personalized scripts that exploit victims' emotional and behavioural vulnerabilities (UNODC 2024a). The rapid rise of AI-generated deception is reflected in the surge of deepfake incidents documented across Australia, Japan, the Philippines, Sri Lanka, and Vietnam from 2022 to 2024 (UNODC 2024a). Deepfakes have become integral to cyber-enabled fraud schemes, particularly sextortion and extortion-based blackmail (UNODC 2024a).

In Cambodia, UNODC has documented cases where trafficked individuals are coerced into participating in video-based scam operations that use AI overlays to simulate nudity during calls, enabling the capture of compromising material to blackmail targets (UNODC 2024a). Voice-cloning technologies have similarly enabled emotionally manipulative scams, including staged kidnappings that combine synthetic voices with publicly available videos to impersonate family members and extort ransom payments (UNODC 2024a).

Analysis of underground marketplaces on the messaging app Telegram reveals a growing commercialization of AI tools marketed explicitly for fraud. Vendors advertise services for large-scale social engineering in fraud schemes, deceptive recruitment, disinformation campaigns, and money laundering (UNODC 2024a). AI has also facilitated biometric forgery, allowing deepfake identity documents to bypass Know Your Customer verification systems (UNODC 2024a). More concerning, criminal groups increasingly jailbreak closed LLMs or manipulate AI input prompts, as well as rely on open-source alternatives, to automate the generation of malicious software capable of evading conventional security controls (UNODC 2024a).

These developments have dramatically reduced the technical threshold for cybercrime. Polymorphic malware—a type of malicious software which is generated and iteratively refined using AI—can alter its code autonomously to evade signature-based detections, placing defenders at a growing disadvantage (UNODC 2025a). As a result, cyber-defence capabilities face mounting pressure from the convergence of scale, speed, and stealth in AI-enabled attacks (UNODC 2025a).

Raising Defences Against AI-Enabled Scams

Paradoxically, AI also underpins many of the most promising countermeasures against cyber fraud. Machine learning, natural language processing, generative AI, and emerging agentic-AI systems enhance real-time fraud detection, compliance monitoring, and cross-institutional intelligence sharing (Rodríguez Valencia et al. 2025). Across Southeast Asia, governments increasingly collaborate with financial institutions, telecommunications providers, and technology platforms to deploy AI-driven systems that move beyond traditional rule-based algorithms.

> “Paradoxically, AI also underpins many of the most promising countermeasures against cyber fraud. Machine learning, natural language processing, generative AI, and emerging agentic AI systems enhance real-time fraud detection, compliance monitoring and cross-institutional intelligence-sharing”

Singapore’s Scam Analytics and Tactical Intervention System illustrates this approach, deploying machine-learning models to detect and block phishing and job-scam websites at scale (GovTech Singapore 2026). As of September 2024, SATIS had detected and blocked 50,000 scam-related websites (GovTech Singapore 2026). Complementing SATIS is the Collaborative Sharing of Money Laundering/Terrorism Financing Information & Cases, a voluntary information-sharing platform developed by the Monetary Authority of Singapore in partnership with leading domestic and international banks to dismantle data silos in anti–money laundering and counterterrorism-financing efforts (Fenergo 2024).

Elsewhere, the Royal Thai Police’s intelligence-sharing partnership with True Corporation reflects growing recognition of the need for public-private coordination in countering cross-border cybercrime. In January 2026, the two entered into a partnership to develop a road map to enhance coordinated operations to disrupt cross-border online crime *(Nation Thailand 2026)*. The partnership is very timely and critical, building on True Corporation’s True CyberSafe, an AI-powered, automated cybersecurity-protection system, and its ongoing crackdown on internet and mobile signals spanning Thailand’s border areas with Cambodia, Laos, and Myanmar (Nation Thailand 2026).

Despite being identified as a key operational hub of scams in Southeast Asia, the Philippines is transforming into an AI-integration hub that leverages human and AI solutions to counter fraudulent activities (Samu 2026). The Business Process Outsourcing (BPO) industry has emerged as an AI-integration hub, combining large-scale human oversight of almost 250,000 personnel with AI-driven solutions for global financial institutions, like JPMorganChase, Deutsche Bank, HSBC, Citi, and PayPal. As an AI-integration hub, the Philippines BPO industry is also pilot testing agentic AI for fraud detection, risk management, and anti–money-laundering compliance (Samu 2026).

A notable regional trend is the deployment of “AI kill switch” mechanisms. Malaysia (Lai 2023), Singapore (Tham 2024), and the Philippines (Bangko Sentral ng Pilipinas 2025) have implemented account-freezing mechanisms that allow suspicious transactions to be halted before funds exit the financial system, signalling a shift from reactive recovery toward preventative intervention.

Implications of AI-Enabled Scams for Regional Co-operation

Despite advances in defensive AI technologies, Southeast Asia remains largely reactive in the face of rapidly adapting transnational crime syndicates. While most ASEAN member states have strengthened domestic frameworks, the region continues to lack a coherent, interoperable mechanism for real-time information sharing and joint enforcement.

Legal disparities, uneven technical capacity, and divergent political will complicate cross-border

coordination (Southeast Asia Public Policy Institute 2024). Even where intelligence sharing occurs, criminal networks continue to relocate, retool, and exploit cryptocurrencies beyond jurisdictional reach. ASEAN acknowledged these shortcomings in its 2025 declaration on combatting cybercrime and online scams, highlighting the urgency of faster asset recovery and coordinated enforcement responses (ASEAN 2025).

Anti-scam measures also introduce economic trade-offs. Stricter verification processes can erode user experience and dampen digital demand. Across the Asia-Pacific, 75 percent of organizations indicate that enhanced fraud controls have reduced customer sales conversion rates (LexisNexis 2023). Implementing novel AI solutions is also becoming an operational burden, imposing additional compliance costs (LexisNexis 2023).

An equally critical piece to addressing cyber scams is the tangible contributions of the tech industry, especially global AI companies. While the Global Fraud Summit in Vienna in March 2026 introduced the Call to Action on Combating Fraud and the Global Public-Private Partnership Framework against Fraud, it is worth asking how much rhetoric turns into action (INTERPOL 2026). A few days before the two-day summit, the Industry Accord Against Online Scams and Fraud was signed by global companies, including Amazon, Google, Meta, Microsoft, and Open AI (Sabin 2026). However, with the adoption of generative AI and the increasing deployment of agentic AI in cyber-scam operations, an important question remains: to what degree should AI companies be held to account to uphold their commitments to such partnerships in measurable and impactful ways?

Demonstrating commitment is an imperative because transparency in the AI industry is rapidly declining. According to the 2025 Foundation Model Transparency Index, the average transparency score out of 100 dropped from 58 in 2024 to 40 in 2025 (Wan et al. 2025). The overall findings show that training data spanning copyright, licences, and Personal Identifiable Information continues to be opaque (Wan et al. 2025). Most companies surveyed withheld basic information about the models they are developing. Amazon, Google, Midjourney, Mistral, OpenAI, and xAI did not provide basic model information indicators encompassing input modality, output modality, model size, model components, and model architecture (Wan et al. 2025).

The current data suggests that while AI developers fear reputational harms from low scores, compelling incentives to uphold transparency is lacking, due to rapid competition with new Chinese AI models like DeepSeek and Alibaba (Wan et al. 2025). With declining transparency, it becomes challenging for tech companies to fully contribute to combatting cyber scams; for example, OpenAI's ChatGPT was found to be widely used in romance scams in fraud compounds (Reuters 2025).

To address the accelerating convergence of AI and cyber scams, Southeast Asia must adopt a more coordinated and forward-leaning approach. First, ASEAN member states should operationalize the 2026 *ASEAN Guide on Anti-Scam Policies and Best Practices* through concrete benchmarks for data sharing, enforcement co-operation, and institutional capacity building.

Governments must demand stronger safeguards from AI-platform developers, including stringent identity verification for Application Programming Interface (API) access—a set of rules that that allow pieces of software to interact with one another—as a standardized mechanism for reporting and mitigating platform abuse.

Finally, increasing transparency within the AI industry is essential. Developers should be required to disclose aggregated information on detected scam-related activities linked to their platforms and co-operate proactively with law-enforcement investigations. Without enforceable transparency, public-private partnerships risk remaining rhetorical as criminal innovation continues to outpace regulation.

REFERENCES

ASEAN (Association of Southeast Asian Nations). 2025. *ASEAN Declaration on Combatting Cybercrime and Online Scams*. September. https://asean.org/wp-content/uploads/2025/09/03.-ASEAN-Declaration-on-Combatting-Cybercrime-and-Online-Scams.pdf.

ASEAN. 2026. *ASEAN Guide on Anti-Scam Policies and Best Practices.* January 9. https://asean.org/wp-content/uploads/2026/01/ASEAN-Guide-on-Anti-Scam-Policies-and-Best-Practices-090126.pdf.

Bangko Sentral ng Pilipinas. 2025. "Circular No. 1213 Series of 2025." December 13. https://www.bsp.gov.ph/Regulations/Issuances/2025/1213.pdf.

Brandolino, John. 2026. "Opening of the Global Fraud Summit." United Nations Office on Drugs and Crime, March 16. https://www.unodc.org/unodc/speeches/2026/160326-global-fraud-summit-opening.html.

Fenergo. 2024. "MAS Launches COSMIC Platform for KYC Data Sharing" (blog). May 7. https://resources.fenergo.com/blogs/mas-cosmic-data-sharing.

Girolamo, Michael Di. 2026. "Deceptive by Design: Social Engineering, Synthetic Media, and the Future of Cyber-Enabled Fraud." C4ADS, February. https://c4ads.org/wp-content/uploads/2026/02/DeceptiveByDesign-C4ADS.pdf.

GovTech Singapore. 2026. "Scam Analytics and Tactical Intervention System." May 12. https://www.tech.gov.sg/products-and-services/for-citizens/scam-prevention/satis/.

GSMA. 2025. "Consumer Trust in Southeast Asia Falters as Cyber Scam Concerns Grow." September 24. https://www.gsma.com/newsroom/press-release/consumer-trust-in-southeast-asia-falters-as-cyber-scam-concerns-grow-new-gsma-commissioned-report-warns/.

INTERPOL. 2026. "INTERPOL–UNODC Global Summit Ends with Call to Action Against Fraud Surge." March 17. https://www.interpol.int/en/News-and-Events/News/2026/INTERPOL-UNODC-global-summit-ends-with-call-to-action-against-fraud-surge.

Lai, Allison. 2023. "Malaysia's Digital Fraud Kill Switch Not a Cure-All, Experts Say." *Asian News Network*, February 27. https://asianews.network/malaysias-digital-fraud-kill-switch-not-a-cure-all-experts/.

LexisNexis. 2023. *The True Cost of Fraud Study.* https://risk.lexisnexis.com/global/en/insights-resources/research/apac-true-cost-of-fraud-study.

Manantan, Mark Bryan. 2023. "Cyber Diplomacy Co-operation on Cybercrime Between Southeast Asia and Commonwealth Countries." *Commonwealth Cybercrime Journal* 133. https://comsec-web-static.s3.eu-west-1.amazonaws.com/s3fs-public/2023-03/D19156-CCJ-1-1-Cyber-Diplomacy-Cybercrime-SE-Asia-Commonwealth--Manantan.pdf.

Nation Thailand. 2026. "AI Adoption and Financial Crime Prevention." February 2. https://www.nationthailand.com/pr-news/pr-news/40062013.

Raman, Jayant, Ashwini Karandikar, Jonas Heckmann, and Sahil Mohnani. 2024. *Cracking the $5 Billion Scam Challenge in Southeast Asia.* Oliver Wyman. https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2024/mar/cracking-$5-billion-scam-challenge-in-southeast-asia-oliver-wyman.pdf.

Reuters. 2025. "ChatGPT Was Used to Help Scammers at an Asia Fraud Compound." *Deccan Herald,* September 15. https://www.deccanherald.com/technology/chatgpt-was-used-to-help-scammers-do-their-thing-at-asia-fraud-compound-2-3728496.

Rodríguez Valencia, Leslie, Maicol Jesús Ochoa Arellano, Santos Andrés Gutiérrez Figueroa, et al. 2025. "A Systematic Review of Artificial Intelligence Applied to Compliance: Fraud Detection in Cryptocurrency Transactions." *Journal of Risk and Financial Management* 18 (11): 612. https://doi.org/10.3390/jrfm18110612.

Sabin, Sam. 2026. "Tech Companies Reach Scam Accord Including Google, Meta, and Amazon." Axios, March 16. https://www.axios.com/2026/03/16/tech-companies-scam-accord-google-meta-amazon.

Samu, Andrew. 2026. "Financial Services Outsourcing in the Philippines: The AI Hybrid Advantage." Disruption Banking, January 14. https://www.disruptionbanking.com/2026/01/14/financial-services-outsourcing-philippines-the-ai-hybrid-advantage/.

Southeast Asia Public Policy Institute. 2024. "Online Fraud in Southeast Asia." March. https://seapublicpolicy.org/wp-content/uploads/2025/09/SEAPPI_Onlne-Fraud_March-2024.pdf.

Tham, Irene. 2024. "What Does a Bank 'Kill Switch' Kill? It Became a Point of Contention After a Victim Lost Money." *Strait Times,* December 16. https://www.straitstimes.com/singapore/what-does-the-kill-switch-kill-it-became-a-point-of-contention-after-a-victim-lost-money.

UNODC (United Nations Office on Drugs and Crime). 2024a. *Transnational Organized Crime and the Convergence of Cyber-Enabled Fraud, Underground Banking and Technological Innovation in Southeast Asia: A Shifting Threat Landscape.* October. https://www.unodc.org/roseap/uploads/documents/Publications/2024/TOC_Convergence_Report_2024.pdf.

UNODC. 2024b. "Billion-Dollar Cyberfraud Industry Expands in Southeast Asia as Criminals Adopt New Technologies." October 7. https://www.unodc.org/roseap/en/2024/10/cyberfraud-industry-expands-southeast-asia/story.html.

UNODC. 2025a. *Emerging Threats: The Intersection of Criminal and Technological Innovation in the Use of Automation and Artificial Intelligence in the Cybercrime Landscape of Southeast Asia.* September. https://www.unodc.org/roseap/uploads/documents/Publications/2025/UNODC_Report_Emerging_threats_-_The_intersection_of_criminal_and_technological_innovation_in_the_use_of_automation_and_AI.pdf.

UNODC. 2025b. *Inflection Point: Global Implications of Scam Centres, Underground Banking and Illicit Online Marketplaces in Southeast Asia.* April. https://www.unodc.org/roseap/uploads/documents/Publications/2025/Inflection_Point_2025.pdf.

Wan, Alexander, Kevin Klyman, Sayash Kapoor, et al. 2025. T*he 2025 Foundation Model Transparency Index.* Stanford Center for Research on Foundation Models. https://crfm.stanford.edu/fmti/December-2025/paper.pdf.

2026

Council for Security Cooperation in the Asia Pacific