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

RSIS COMMENTARY

www.rsis.edu.sg No. 130 – 6 September 2024

Enhancing ASEAN Cooperation Against AI-Powered Cyber Influence Operations

By Benjamin Ang and Muhammad Faizal

SYNOPSIS

Artificial intelligence is increasingly used to power cyber influence operations in geopolitical competitions and conflicts, especially after the acceleration of generative AI development in 2022. As such operations could impact the security of countries in Southeast Asia, the ASEAN member states should study them in depth and cooperate more in information sharing.

COMMENTARY

OpenAI, the organisation that developed the artificial intelligence (AI) tool ChatGPT, announced in August 2024 that it had identified and banned accounts using the tool to generate content for an influence operation targeting the 2024 US presidential campaign. With help from Microsoft's threat intelligence reports, OpenAI found that this operation was likely to be linked to Iran. The operation used ChatGPT to create socially polarising articles to share on social media and websites.

Similar to how commercially available drone technology has made airpower more attainable by lesser military actors, commercially available generative AI, which is AI capable of generating text, images, videos, or other data using generative models, places more information power in the hands of such actors engaged in conflicts with major cyber powers.

AI is increasingly being applied in cyber influence operations as part of geopolitical competitions or information warfare during armed conflicts. Its utility has enhanced the strategic importance of the digital battlespace, increasing the risk of conflict involving the major powers.

Generative AI and the Digital Battlefield

This is not the first time AI has been used in influence operations. Earlier this year, Microsoft and OpenAI disrupted the operations of threat actors ostensibly connected to Russia, Iran, China, and North Korea, who used OpenAI services like ChatGPT not only in influence operations but also to create content for use in phishing campaigns, research the ways that processes could be hidden in a system, and the ways malware could evade detection. OpenAI also revealed that Israel had used ChatGPT to produce deceptive content that praised Israel's conduct in the war in Gaza.

Ominously, threat actors – particularly the North Korea-linked threat actor Emerald Sleet – have used AI tools to identify experts and organisations dealing with defence issues in the Asia-Pacific. Besides using AI with autonomous systems to identify human targets for elimination, the technology can be used to identify targets for influence operations or intelligence collection through spear-phishing attacks.

Earlier, cyber influence operations also attempted to use AI to create socially polarising content, especially around armed conflicts like the Russia-Ukraine war and the Israeli-Hamas war. Such operations aimed to demoralise enemy troops and destabilise the opponent's society. For example, in 2022, following the Russian invasion of Ukraine, threat actors circulated deepfake videos falsely depicting Ukrainian President Volodymyr Zelensky and his military leaders spreading messages aimed at confusing the Ukrainian public and damaging troop morale.

Iran's use of ChatGPT to influence the 2024 US presidential campaign is one of the latest cases illustrating the evolving and disruptive risks to national and regional security that come with rapid advancements in AI technology. The technology's prowess in geopolitical and military conflicts was less pronounced before the sudden acceleration in generative AI development around 2022.

The technological tools available today are still insufficient for detecting deepfakes where generative AI is used. Moreover, the increasing distrust due to the use of deepfakes and lack of digital literacy could cause people to misperceive *bona fide* content as deepfakes.

A Concern for ASEAN Security?

Why is the use of generative AI in Iranian cyber influence operations targeting the US presidential campaign an issue for ASEAN member states to be concerned about?

At the strategic level, AI-powered cyber influence operations targeting the US presidential campaign would widen schisms in American society and potentially impact US foreign policy and defence strategy. This could cause major powers worldwide to adjust their security postures, a scenario that would create implications for Asia-Pacific security, including that of ASEAN.

At the operational level, Iranian cyber influence operations illustrate the transnational nature of cyber and information threats. Malicious but persuasive content can be created in one place and then digitally transmitted around the world. Threat actors could use deceptive "phishing" emails to trick victims across borders into downloading

malware or disclosing passwords, enabling them to launch cyber or information attacks thousands of kilometres away.

Such threats to national and regional security are complex and can only be fended off through cooperation between states. States will respond better to such threats if there is information exchange regarding influence operations, whether AI-powered or not. States with fewer resources and experience in dealing with such threats could also benefit from capacity-building programmes by the digitally advanced states, which would, in turn, benefit from the uplifted cyber resilience and digital literacy of their partners.

Enhancing Digital Defence Cooperation

Unlike the civilian sector, information sharing and capacity building are more challenging in the defence sector, where threat information and intelligence are often closely guarded secrets. But now that AI-powered cyber influence operations are playing out in geopolitical competition and armed conflicts, the defence establishments of ASEAN member states should leverage existing platforms to enhance cooperation.

In this regard, the defence establishments of ASEAN member states should maximise the potential of the Malware Information Sharing Platform established by the ADMM Cybersecurity and Information Centre of Excellence (ACICE) and use the opportunities offered by the Digital Defence Symposium that the latter co-organises to examine AI-powered cyber influence threats in greater depth along with the civilian sector and tech companies.

Senior Fellow Benjamin Ang is the Head of the Centre of Excellence for National Security (CENS) and Head of Digital Impact Research (DIR) at the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore. Muhammad Faizal is a Research Fellow in the Regional Security Architecture Programme at RSIS' Institute of Defence and Strategic Studies (IDSS).

S. Rajaratnam School of International Studies, NTU Singapore
Block S4, Level B3, 50 Nanyang Avenue, Singapore 639798
T: +65 6790 6982 | E: rsispublications@ntu.edu.sg | W: www.rsis.edu.sg