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

Science, Technology and Security

January 2026

Published by the Future Issues and Technology (FIT) Research Cluster, RSIS. This Bulletin comes as a series of articles on science and technology from the angle of national security.

Responding to AI-Enabled Cybercrime: Governance, Attribution, and Escalation | Helena Yixin Huang

Drawing on insights from a tabletop exercise conducted in Singapore in October 2025, this article examines decision-making, governance, and escalation challenges associated with artificial intelligence (AI)-enabled cybercrime, and considers their implications for policymakers across both the public and private sectors.

Many discussions on AI-enabled cybercrime have focused on its novelty, particularly on how AI serves as a tool for new attack techniques and new categories of risks. Yet, as AI becomes increasingly embedded across cybercrime operations, the most consequential shift may lie elsewhere. AI has not fundamentally altered cybercriminals' motivations, nor has it rendered existing cybersecurity principles obsolete. Instead, by accelerating the speed of incidents, it has compressed decision-making timelines and intensified the pressure on governance, coordination, and judgement.

These dynamics were explored during a tabletop exercise held in Singapore in October 2025, organised in collaboration with the University of California, Berkeley's Center for Long-Term Cybersecurity and cybersecurity company Fortinet. The exercise brought together participants from government agencies, critical infrastructure operators, private sector companies, and academia. It examined a series of AI-enabled cyber incidents, beginning with a corporate breach and escalating to disruptions across essential services.

While the exercise was framed around AI-enabled cybercrime, the discussions at various points extended beyond criminal activity, highlighting how early-stage cyber incidents often unfold before intent or attribution is clear. Ultimately, the exercise illuminated how institutions must adapt their response strategies to the unprecedented speed and uncertainty introduced by AI..

Observation 1: Organisational clarity matters as much as technical capability in early response

Before discussions at the tabletop exercise turned to tools or forensic processes, participants prioritised clarifying roles, authority, and decision-making responsibilities. This did not reflect technical shortcomings; rather it highlighted how shared

understanding becomes critical when responding under time constraints. This dynamic was particularly evident given the participants' diverse institutional backgrounds, with each bringing implicit assumptions about escalation, reporting, and communication. In the absence of immediate clarity, participants' attention was divided between action and interpretation, which can quickly erode the coordination necessary for effective early response. Technical preparedness alone is therefore insufficient.

Observation 2: Attribution is a strategic policy choice, not a technical endpoint

When the question of attribution arose, participants approached it with notable restraint. Rather than treating attribution as the conclusion of technical investigations, they framed it as a strategic policy choice shaped by legal, diplomatic, and public-trust considerations. Containment, stabilisation, and confidence management were consistently emphasised over the act of naming an adversary. Participants raised concerns about false flags, noting that attackers may deliberately reuse Tactics, Techniques, and Procedures (TTPs) to mislead investigators or divert suspicion. Consequently, attribution was seen as a judgement that must account not only for technical evidence, but also for uncertainty, intent, and potential geopolitical consequences, rather than as a routine endpoint of investigation.

Observation 3: AI complicates escalation judgements by narrowing the margin of certainty

Throughout the exercise, escalation decisions, such as widening internal decision-making and engaging external entities, remained deliberate and threshold-driven. Participants perceived the early indicators of compromise or anomalous system behaviour as insufficient grounds for immediate escalation and coordination. Escalation is typically triggered when indicators accumulate, such as increasing evidence of shared infrastructure being attacked, or when the impact is determined to extend beyond a single organisation or sector. At the same time, the involvement of AI complicates these judgements. Participants noted that AI-enabled malware and automation could enable incidents to scale or adopt more quickly, thus narrowing the margin of confidence for timely escalation.

Policy Implications

The observations from the tabletop exercise reveal policy challenges that transcend technical capability or tool deployment. AI-enabled cybercrime places strain on governance structures, decision-making processes, and coordination mechanisms that were largely designed for slower-moving and more clearly defined incidents. Addressing these pressures will require a shift in how institutions organise authority, manage uncertainty,and coordinate responses under conditions where speed and ambiguity coexist.

First, organisational clarity must be elevated from an operational detail to a core pillar of cyber preparedness. AI-enabled cybercrime compresses response timelines and increases uncertainty, leaving less room for deliberation once the clock starts ticking. Ambiguity around decision and escalation authority, as well as communication responsibilities, can create bottlenecks as damaging as technical gaps. Policy frameworks and incident response doctrines should therefore explicitly clarify the

governance structures. The arrangements also need to be clearly articulated and understood across the different team functions. Without such clarity, well-managed resourced systems can risk losing time at critical moments.

Second, attribution should be governed by clear policy guidelines that distinguish technical investigation from strategic decision-making. While technical processes can generate TTPs, attribution is in itself a judgement of disclosure, signalling, and consequences. Policies should therefore establish thresholds for when attribution is necessary versus when it remains optional, and include guidance on how uncertainty should be handled internally. Treating attribution as a policy choice rather than an endpoint to a technical response and investigation can help prevent both overreaction and reaction paralysis, thereby preserving credibility and trust with both internal stakeholders and external partners.

Third, escalation frameworks should be recalibrated to support graduated coordination rather than binary responses. Existing escalation models often assume a linear progression from detection, confirmation, to crisis, with limited space for other intermediate steps. Flexible escalation pathways allowing early information sharing, sector-level coordination, or even government engagement without automatically triggering public disclosures or emergency measures would be useful. The objective is not to escalate incidents earlier or faster, but to allow coordination to scale proportionally with the emerging scope and impact of the incident.

Taken together, these policy implications point to a broader conclusion: tackling AI-enabled cybercrime does not require entirely new policy architectures. Instead, it needs existing policies to be sharpened across both the public and private sectors. Governance structures, attribution practices, and escalation pathways must be designed to function under the conditions brought about by AI, namely speed, uncertainty, and ambiguity. Treating these components of response and coordination as interconnected policy challenges will be critical to maintaining effective and credible responses in an increasingly AI-mediated threat environment.

About the Author

Helena Yixin Huang is an Associate Research Fellow at the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University, Singapore. Her research examines cybercrime, cybersecurity policies, and the governance of emerging technologies, with a particular focus on how evolving forms of digital criminality shape international and regional responses. Helena also collaborates with international organisations such as INTERPOL, the United Nations Office on Drugs and Crime, and the Global Initiative Against Transnational Organized Crime to strengthen multistakeholder approaches to combating cybercrime.

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