

<Appendix: About the Trilateral Study Group on AI, Human Control, and Strategic Stability>

Overview

The Trilateral Commission (Secretariat: International House of Japan) has launched the Trilateral Study Group on AI, Human Control, and Strategic Stability. The Study Group brings together expertise from the three regions of the Trilateral Commission (Asia-Pacific, North America, and Europe) to examine the impact of AI on security, and in particular on strategic stability, including nuclear deterrence. It will address a core challenge for the AI age: how to preserve meaningful human control, accountability, and responsibility in decisions involving force, escalation, and strategic stability.

Background and Rationale

Artificial intelligence is becoming one of the defining forces of our age, altering the relationship between human judgment, industrial society, and political power. Its implications are especially consequential for security and strategic stability. AI is already shaping intelligence analysis, cyber defense, logistics, autonomous systems, critical infrastructure protection, targeting support, decision support, and command-and-control. The same attributes that make AI powerful (speed, scale, pattern recognition, and adaptive decision support) can also compress decision time, obscure accountability, and increase the risk of inadvertent escalation. Used responsibly, AI can strengthen resilience, legal compliance, and crisis management; however when used without adequate safeguards, risks may materialize.

AI in security and defense is entering a governance space that existing frameworks do not yet adequately address. The debate on lethal autonomous weapons remains essential, yet captures only one part of the problem. Civilian AI governance addresses safety, transparency, privacy, and innovation, yet leaves unresolved the strategic consequences of AI as it becomes embedded in the institutions, infrastructures, and decision processes through which states exercise power. These questions remain insufficiently explored in existing international discussions, and the Study Group will address this gap.

Why the Trilateral Commission

The Trilateral Commission is not a treaty forum, a military alliance, or a technical standards body. Its value lies in creating a candid space where leaders from politics, business, technology, law, academia, and security communities can think together before crises harden into division. The Study Group draws on several strengths distinctive to the Commission.

First, the Commission connects strategic, technological, legal, industrial, and political perspectives, and links the Indo-Pacific and Euro-Atlantic theatres. It offers a flexible, lightweight platform that can help shape a responsible path between an unrestricted AI race or an unrealistic development freeze.

Second, through the founding networks it has cultivated in finance, the Commission brings to the economic dimensions of AI security a vantage point that security-focused forums lack. Issues such as capital flows around AI, government procurement, and shifts in industrial structure shape how AI capabilities spread and who controls them. These questions sit at the core of the Commission's historic strengths.

Third, the Commission is non-partisan and can convene experts across political beliefs, preserving a continuity of expertise that no government process currently sustains.

Fourth, the Commission maintains points of contact with diverse policy and technical communities across regions, and can sustain dialogue informed by different political and social contexts. The Asia-Pacific region in particular can contribute perspectives largely absent from existing processes, including diverse intellectual traditions on ethics and human responsibility.

Focus of the Study

The Study Group will examine, across security and defense applications, how AI-enabled systems are changing the conditions under which human beings exercise judgment, assign responsibility, and manage risk. Particular attention will be given to areas where speed, autonomy, uncertainty, and strategic consequence converge, and the work centers on the following lines of inquiry.

The first focus will be concerning the role of AI in nuclear command-and-control. The Study Group will examine the implications of AI in areas of the highest consequence, including nuclear command-and-control support, strategic early warning, escalation assessment, and decision support. Here AI acts primarily as a factor that amplifies known risks: faster targeting compresses decision time, and decision-support tools can foster unacknowledged dependence. The Study Group will seek to identify where the use of AI may be unacceptable,

where strict human control is indispensable, and where additional safeguards are necessary. A normative understanding on AI and nuclear weapons is widely regarded as attainable, and is a rare area of near-universal concern.

Second, the intersection of AI and biosecurity. The convergence of AI and the life sciences raises risks of comparable consequence, as AI-enabled design and synthesis can lower the barriers to developing dangerous agents, including by non-state actors. This is also an area in which all three regions are actively engaged, and the work will proceed in coordination with related efforts, such as work toward verification arrangements under the Biological Weapons Convention.

Third, the interaction between different AI systems. The question of what happens when forces and decision systems built on AI models grounded in different data, values, and assumptions confront one another in a crisis is one that no existing international process addresses. When such systems interact at machine speed, cultural and strategic assumptions may diverge invisibly, and escalation may proceed in ways neither side intends or understands. The Study Group will develop the concept of shared expectations, communication practices, and restraint measures that would apply when AI-enabled forces built on different models interact.

Running through these lines of inquiry, the Study Group will also consider how responsibility is maintained across the life cycles of autonomous systems, how institutions preserve time, discipline, and human judgment under machine-speed conditions, and how high-level principles are translated into practical guidance for operators, policymakers, technologists, lawyers, and political leaders.

Furthermore, the Study Group aims to complement the ongoing initiatives of international organizations and governments—including the United Nations, the Convention on Certain Conventional Weapons (CCW), the Summit on Responsible Artificial Intelligence in the Military Domain (REAIM), NATO, the European Union, and individual national governments—while bridging the gap between high-level principles and practical guidelines. In particular, our work will proceed in alignment with key upcoming developments, notably the deliberations at the First Committee of the UN General Assembly in October 2026, the Working Group on the Biological Weapons Convention (BWC), and the UN permanent process on cybersecurity.

Process and Output of the Study

The Study Group is led by Co-Chairs designated by each of the three regions, supported by a small group of experts. Its main output will be a concise issue paper on AI, human control, and strategic stability, identifying key risks, emerging areas of consensus, unresolved questions, and practical options for discussion. Preliminary findings will be shared through the three regional meetings held from autumn 2026 onward, and, following further discussion, the Study Group aims to prepare a set of recommendations in 2027. The organizational structure for this initiative is outlined below:

	Asia-Pacific	Europe	North America
Co-Chairs	Ken Jimbo Managing Director, International House of Japan ; Professor, Keio University	Roberto Cingolani Former CEO, Leonardo	Doug Beck Board Member, Center for a New American Security
Advisors	Yun Byung-se Chairman, Seoul International Law Academy (SILA); former Chair, the Global Commission on Responsible AI in the Military Domain (GC REAIM); former Minister of Foreign Affairs, Republic of Korea Archish Mittal Special Advisor, Iron Pillar Fund	Marietje Schaake Non-resident Fellow, Stanford University; former Member of the European Parliament Jacobo Roa Vicens Machine Learning Centre of Excellence, JPMorgan Chase & Co.; Centre for Computational Statistics and Machine Learning, University College London	Tom Jenkins Chair of the Board, OpenText Corporation Michael B. Greenwald Global Head of Financial Innovation and Digital Assets, Amazon Web Services

(Reference) Role of the Asia-Pacific Group

The Asia-Pacific Group will provide perspectives that have not been sufficiently represented in existing international discussions. Specifically, these include East Asian philosophical traditions concerning ethics and human responsibility; the ethical implications for the civilian sector amidst the shift from "dual-use" to "mixed-use," where the boundaries between military and civilian applications become increasingly blurred; and research trends aimed at reintroducing cultural and contextual values into AI systems developed under singular value frameworks.