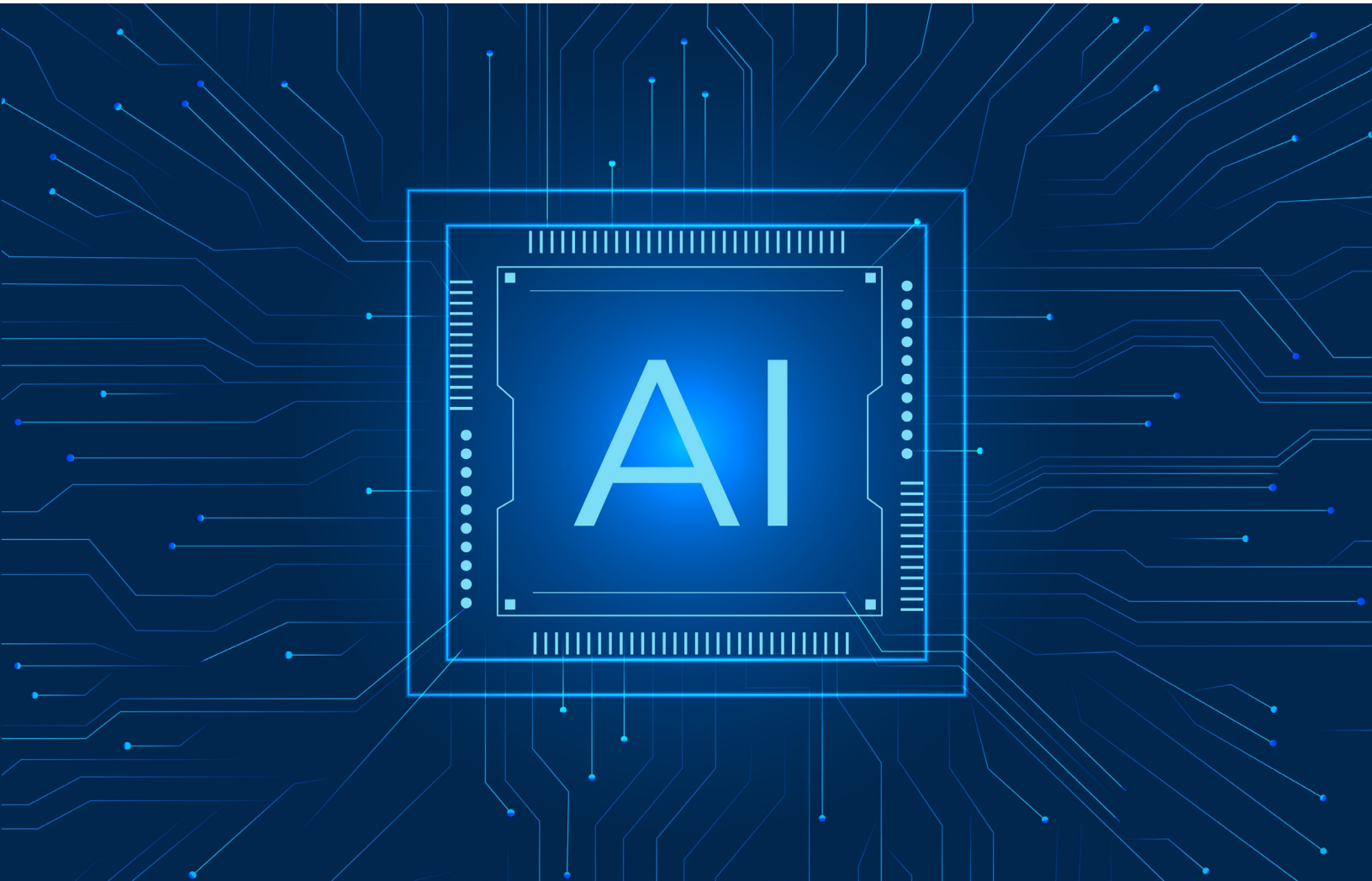
SECURITY COUNCIL REPORT
Independent. Impartial. Informative.

Policy Paper | September 2026

ARTIFICIAL INTELLIGENCE, GEOPOLITICS, AND THE UN SECURITY COUNCIL

Cover illustration: © Crafticstudio/Shutterstock



Contents

Security Council Report would like to thank the government of Greece for its generous support of this project. We also express our gratitude to all colleagues who generously shared their perspectives on the topics and issues covered in this report.

Introduction

Artificial intelligence (AI) is becoming increasingly relevant to the work of the Security Council, with the potential both to support and to undermine international peace and security. It can bolster efforts to prevent, manage, and resolve conflict, including by helping to identify patterns of violence, monitor ceasefires, and strengthen UN peacekeeping and mediation.[1] Its growing capabilities also present a range of security risks. As a dual-use technology with applications across a wide range of domains, AI can lower barriers to malicious cyber activity targeting critical infrastructure, including maritime infrastructure, and potentially facilitate the development of biological weapons. It can also make it easier to produce and disseminate convincing misinformation and disinformation at scale, and affect the conduct of armed conflict, including by compressing decision-making timelines and enabling increasingly autonomous weapons systems.[2]

More broadly, AI is reshaping geopolitical competition. As advanced AI systems become more capable, demand for semiconductors, data, and energy is increasing. Access to these resources and the supply chains that support them is becoming a source of strategic competition and potential geopolitical friction.[3]

As AI capabilities advance, existing risks may grow in scale and complexity and new risks may emerge, warranting greater attention from the Council. The Council's engagement on AI as a standalone thematic issue has so far been sporadic and largely dependent on the priorities of individual Council members. Its mainstreaming of AI-related risks into existing thematic and country-specific agenda items also remains inconsistent. A more systematic approach is needed to strengthen the Council's ability to identify emerging risks, assess their implications for international peace and security, and consider appropriate responses.

While Council dynamics, particularly the divergent views on the Council's role in addressing AI, are likely to limit prospects for agreement on new mandates or formal mechanisms, members have several options for strengthening engagement within existing working methods and practices.

This policy paper examines how the Security Council can strengthen its engagement on AI to better address the technology's peace and security implications. Section 1 considers AI's geopolitical effects, including competition over capabilities, resources, and governance. Section 2 examines its implications for international peace and security, including both risks and opportunities. Section 3 assesses the Council's engagement to date and related gaps, while Section 4 provides options for more systematic and anticipatory engagement. The paper uses the maritime domain to illustrate how different AI-related risks and opportunities can affect maritime security and issues already before the Council.

WHAT IS ARTIFICIAL INTELLIGENCE?

There is no universally accepted definition of AI. The term generally refers to computational systems that can perform tasks such as learning, reasoning, prediction, problem-solving, and decision-making. Machine learning is a branch of AI in which algorithms learn patterns from data rather than relying solely on explicitly programmed rules, while deep learning uses multi-layered artificial neural networks to process complex forms of data, including text and images.[4]

Recent attention has focused on generative AI and large language models (LLMs). Generative AI systems can produce content such as text, images, audio, video, and code, while LLMs are trained on large volumes of text data to generate and process language. AI agents use external tools to perform multistep tasks and pursue specified goals with varying degrees of human oversight.5

Frontier AI generally refers to the most capable general-purpose AI systems. Their development has been shaped by empirical scaling laws, under which performance has tended to improve with greater computational resources, training data, and model scale. Continued scaling remains an important approach to developing more capable systems, although how far it can sustain further advances remains uncertain.

AI systems can be accessed in different ways. Many are available through developer-controlled interfaces, allowing companies to retain control over access and apply safety measures. Others are released as "open-weight" models, allowing users to download and modify the model's parameters and run the model on their own hardware. Such models can support competition and allow countries and institutions to operate AI systems without relying on continued access through foreign providers. At the same time, users can remove safeguards and refusal mechanisms built into the model.

Discussions of more advanced AI often distinguish current systems from artificial general intelligence (AGI) and artificial superintelligence (ASI). Definitions of both terms vary. AGI can refer to systems capable of matching or exceeding human performance across a broad range of cognitive tasks, or of outperforming humans at most economically valuable work, while ASI refers to hypothetical systems whose capabilities substantially exceed those of humans across a broad range of cognitive domains, potentially surpassing the collective capabilities of large groups or organisations.[6] It is unclear whether such systems will be developed and, if so, when. Nevertheless, the possibility of their development is the subject of serious consideration among leading AI researchers and developers.[7]

1 Secretary-General, "Remarks to the Security Council on Artificial Intelligence", 18 July 2023.
2 International AI Safety Report 2026; A/80/78 (5 June 2025).
3 UN Trade and Development (UNCTAD), Technology and Innovation Report 2025: Inclusive Artificial Intelligence for Development; UNCTAD, World Investment Report 2026: International Investment in a Turbulent Era, 2026.
4 Organisation for Economic Co-operation and Development (OECD), "Explanatory Memorandum on the Updated OECD Definition of an AI System", OECD Artificial Intelligence Papers No. 8 (2024); International AI Safety Report 2026.
5 International AI Safety Report 2026.
6 OECD, Introducing the OECD AI Capability Indicators (2025); Morris et al., "Levels of AGI for Operationalizing Progress on the Path to AGI", ICML 2024; Genewein et al., "From AGI to ASI" (Google DeepMind, 2026).
7 Grace et al., "Thousands of AI Authors on the Future of AI," Journal of Artificial Intelligence Research 84 (2025).

Section 1: The Geopolitics of AI

AI is increasingly underpinning countries' economic, military, and scientific capabilities, and their diplomatic influence. The advantages associated with leadership in AI development and deployment have contributed to intensifying international competition over AI, prompting comparisons with an arms race.[8] This competition extends beyond the development of frontier models. It also encompasses access to the physical and technological inputs required to develop and deploy AI, as well as the capacity to influence emerging technical standards and governance frameworks.

Competition is further being driven by expectations about where the technology may lead. Governments and companies are investing heavily in increasingly capable systems, including in pursuit of AGI or other forms of advanced AI that could confer substantial strategic advantages. Whether such systems can be developed and whether developing them first would provide a decisive or durable advantage remain unclear. Nevertheless, the prospect of such an advantage can itself shape state behaviour and encourage actors to accelerate investment and intensify competition.[9]

Regardless of how this competition evolves, some of its clearest implications for international peace and security concern the material foundations of AI. Access to semiconductors, computing infrastructure, data, energy, and critical minerals can create dependencies and exposure to supply disruption. These concerns intersect with many existing issues on the Security Council's agenda and are therefore particularly relevant to its work.[10]

Competition over AI Resources and Infrastructure

Advances in AI depend on an ecosystem encompassing advanced semiconductors and their supply chains, data, data centres and cloud-computing infrastructure, critical minerals, and reliable access to electricity and water. These resources and capabilities are unevenly distributed among member states. Many developing countries possess significant critical-mineral reserves and production capacity, while advanced computing and data infrastructure remain concentrated in a relatively small number of technologically advanced economies.[11]

Advanced semiconductors have become a potential point of geopolitical leverage. The production of leading-edge chips and key semiconductor manufacturing equipment remains highly concentrated, while export controls have increasingly been used as instruments of strategic competition to restrict access to critical technologies.[12]

Controls have also begun to extend from hardware to AI models themselves. In June 2026, the US Department of Commerce directed Anthropic to suspend access by foreign nationals to two newly released frontier models on national security grounds, prompting the company to suspend them for all customers before restrictions began to be lifted later that month.[13] The episode illustrates how dependence on foreign providers can expose users in other jurisdictions to restrictions imposed by the provider's home government.

Access to large and diverse datasets has similarly become strategically important for AI development and deployment, as the composition and quality of training data can influence system performance and outputs.[14] Concerns about control and jurisdiction have prompted some governments to require certain data to be stored or processed domestically.[15] Such measures form part of a wider push towards "sovereign AI", as states seek greater domestic control over computing capacity, data, infrastructure, and models.[16]

The internationalised nature of AI supply chains and widespread reliance on foreign cloud providers make full domestic control difficult.[17] Greater technological autonomy may therefore depend on reducing reliance on any single external provider or jurisdiction and maintaining credible alternatives. This could involve a form of multi-alignment in AI, whereby states diversify their access to models, computing infrastructure, and hardware and, where appropriate, pool capabilities with other states.

These realities also have implications for international peace and security. Rising demand for critical minerals and other resources required for advanced technologies could intensify competition in resource-rich countries affected by conflict, weak governance, or illicit resource extraction. In such settings, revenues from illicit resource exploitation can strengthen armed and criminal actors, while disputes over access, ownership, territorial control, and revenue-sharing can shape parties' interests and incentives for continued conflict.[18] Competition to secure critical-mineral supply chains is also extending beyond extraction to transport corridors and other enabling infrastructure, creating additional arenas for geopolitical competition.[19] The material requirements of AI can therefore interact with existing sources of instability and geopolitical rivalry.

Unequal access to advanced AI can also reinforce existing technological and economic disparities. Many developing countries have limited access to the computing capacity and other resources needed to develop and deploy advanced systems, which can limit their ability to capture higher-value economic benefits, adapt AI systems to local priorities, and influence emerging standards and governance frameworks.[20] These asymmetries also have security implications. Unequal access to advanced military AI can widen capability gaps between states, while the wider diffusion of increasingly capable systems could increase access for terrorist groups and other malicious non-state actors.[21]

8 UNDP, Human Development Report 2025; Secretary-General, "Remarks to the Security Council on Artificial Intelligence", 19 December 2024.
9 Mitre et al., eds., The Artificial General Intelligence Race and International Security (RAND, 2025); International AI Safety Report 2026.
10 UNCTAD, Technology and Innovation Report 2025; OECD, Competition in Artificial Intelligence Infrastructure, no. 330 (2025).
11 UNCTAD, Technology and Innovation Report 2025.
12 OECD, Mapping the Semiconductor Value Chain, No. 182 (2025); OECD, Economic Security in a Changing World (2025).
13 Anthropic, "Claude Fable 5 and Claude Mythos 5", 9 June 2026; Axios, "Scoop: Trump admin blocks foreign access to Anthropic's most powerful AI", 12 June 2026; CNBC, "Anthropic Says Trump Admin Has Lifted Export Controls on Claude Fable 5 and Mythos 5", 30 June 2026; Al Jazeera, "US Lifts Restrictions on Anthropic's Powerful AI Models Fable and Mythos", 1 July 2026; Forbes, "Anthropic Disabled Fable 5 and Mythos 5 After a U.S. Export-Control Order", 16 June 2026.
14 OECD, "Mapping Relevant Data Collection Mechanisms for AI Training", AI Papers No. 48 (2025); André et al., "Developments in Artificial Intelligence Markets," AI Papers No. 37 (OECD, 2025).
15 OECD/WTO, Economic Implications of Data Regulation: Balancing Openness and Trust (2025).
16 International Telecommunication Union (ITU), The Annual AI Governance Report 2025: Steering the Future of AI.
17 OECD, Economic Security in a Changing World; UNCTAD, Technology and Innovation Report 2025.
18 Secretary-General, remarks to the Security Council open debate on natural resource governance, 22 July 2026.
19 Vandome and Dideberg, "The Lobito Corridor Shows How Digitization Is Shaping Critical Minerals Competition", Chatham House, 20 July 2026; Clyde Russell, "Can Africa Win as the West and China Scramble for Minerals?" Reuters, 12 February 2026.
20 UNCTAD, Technology and Innovation Report 2025; ITU, Annual AI Governance Report 2025.
21 A/80/78 (2025).

AI SUPPLY CHAINS AND MARITIME INFRASTRUCTURE

Competition over AI inputs intersects with maritime security, as many of the physical resources and equipment needed for AI development pass through shipping routes and ports, while subsea infrastructure carries much of the data and digital connectivity on which internationally distributed AI systems depend.

More than 80 percent of international trade in goods by volume is carried by sea, including many of the minerals, semiconductor inputs, computing equipment, and energy resources on which AI development depends.[22] These flows are exposed to disruption at maritime chokepoints. Where AI-related production or processing is already concentrated, disruption to maritime transport could compound existing supply constraints.

Ports are another point of vulnerability. They connect maritime transport with inland logistics and manufacturing and can therefore affect how quickly international inputs reach semiconductor facilities, data centres, and other parts of the AI supply chain. Ports have also become increasingly dependent on digital systems for terminal management and cargo processing, creating exposure to cyber disruption.[23] AI-related supply chains therefore depend partly on the physical and digital resilience of the ports through which relevant materials and equipment pass.

These risks also extend beneath the sea. Submarine telecommunications cables carry more than 99 percent of international data traffic and support cloud computing and other cross-border digital services.[24] Because cloud infrastructure and cross-border data flows form part of the wider infrastructure on which AI development and deployment depend, damage to subsea cables can disrupt another layer of that infrastructure. Cables remain exposed to accidental damage from anchoring and fishing as well as natural hazards. Since 2023, a series of cable-damage incidents has also increased concern about possible intentional damage.[25]

Regulatory Divergence and Multilateral Governance

Geopolitical competition can make multilateral governance more challenging, as states may be reluctant to accept rules they believe could constrain technological development or reduce their relative advantage.[26] Approaches to AI governance reflect different national policy priorities and perceptions of risk. While states may agree on broad principles for responsible AI, they often differ over how those principles should be translated into rules and enforced in practice.

The EU has adopted a comprehensive and legally binding framework that applies different requirements according to the risks posed by AI systems, alongside specific obligations for general-purpose models.[27] The US does not have an equivalent federal AI law and has relied on existing sectoral laws and regulations, state-level measures, technical standards, and voluntary risk-management frameworks.[28] China, meanwhile, combines strong state direction of AI development with binding rules targeting specific applications and risks, including algorithmic recommendation systems, generative AI, and AI-generated content.[29] Singapore, for example, introduced guidance specifically addressing the responsible deployment of agentic AI in January 2026.[30]

At the international level, AI governance is developing through multiple institutions and processes. Within the UN system, member states adopted the Global Digital Compact in September 2024, setting out broad commitments on AI governance.[31] The General Assembly subsequently established the Global Dialogue on AI Governance and the Independent International Scientific Panel on AI in August 2025.[32] The panel issued its preliminary report on 1 July 2026, and the first session of the Global Dialogue was held on 6 and 7 July.[33] Other UN processes address related dimensions of AI governance and use, including UNESCO's work on AI ethics, General Assembly processes on AI in the military domain, and the Group of Governmental Experts (GGE) on emerging technologies in the area of lethal autonomous weapons systems under the Convention on Certain Conventional Weapons (CCW).[34] The permanent Global Mechanism on information and communications technologies (ICTs) in the context of international security also considers the effects of AI on the ICT threat environment.[35]

Beyond the UN, the OECD, G7, and G20 have developed AI principles and policy frameworks, while international AI summits and the Responsible AI in the Military Domain (REAIM) process have addressed specific governance and security questions.[36] Regional approaches include the AU's Continental Artificial Intelligence Strategy, endorsed in 2024, and ASEAN's Guide on AI Governance and Ethics, issued the same year and subsequently expanded to address generative AI.[37] In July 2026, 29 countries signed an agreement establishing the World Artificial Intelligence Cooperation Organization (WAICO), an independent intergovernmental organisation headquartered in Shanghai and intended to promote international

22 UNCTAD, Review of Maritime Transport 2021.
23 International Maritime Organization (IMO), Guidelines on Maritime Cyber Risk Management (2025); International Association of Ports and Harbors, Cybersecurity Guidelines for Ports and Port Facilities, Version 1.0 (2021).
24 ITU, Global Connectivity Report 2025.
25 Insikt Group, Submarine Cables Face Increasing Threats Amid Geopolitical Tensions and Limited Repair Capacity (Recorded Future, 2025).
26 UN High-level Advisory Body on AI, Governing AI for Humanity: Final Report (2024).
27 European Commission, "AI Act".
28 The White House, "Ensuring a National Policy Framework for Artificial Intelligence", Executive Order 14365 (11 December 2025).
29 Harris, Regulating Artificial Intelligence (Congressional Research Service, 2025).
30 Singapore IMDA, "Singapore Launches New Model AI Governance Framework for Agentic AI", 22 January 2026.
31 UN General Assembly, Pact for the Future, A/RES/79/1 (22 September 2024), Annex I.
32 A/RES/79/325 (26 August 2025).
33 A/RES/79/325 (2025); Independent International Scientific Panel on AI, Preliminary Report (2026); UN, "Global Dialogue on AI Governance", 6–7 July 2026.
34 UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021); UN General Assembly, A/RES/79/239 (31 December 2024) and A/RES/80/58 (5 December 2025).
35 UN Office for Disarmament Affairs (UNODA), "Global Mechanism on ICTs in the Context of International Security – Plenary (2026)", 20–24 July 2026.
36 OECD, Recommendation on Artificial Intelligence, OECD/LEGAL/0449 (2019, amended 2024); G7, Hiroshima AI Process Comprehensive Policy Framework (2023); UNCTAD, Technology and Innovation Report 2025; Government of the Netherlands, "Call to Action on Responsible Use of AI in the Military Domain", 16 February 2023.
37 AU, Continental Artificial Intelligence Strategy (2024); ASEAN, ASEAN Guide on AI Governance and Ethics (2024) and its expanded guidance on generative AI (2025).

cooperation and global AI governance. Its founding members include Brazil, China, Indonesia, Russia, and South Africa.[38]

These efforts can complement one another, but differences in membership and mandates can also produce duplication and uneven representation. According to UNCTAD, 118 countries, primarily in the Global South, were not parties to any of the seven international AI governance initiatives or instruments it examined.[39] Nevertheless, across these initiatives, areas of convergence have emerged around safety, transparency, and accountability, alongside a broader shift towards risk-based governance.[40] The Security Council, which is beginning to define its own role on AI in relation to its mandate, must therefore navigate an increasingly crowded field and ensure that its work complements these processes rather than duplicating them.

Section 2: How AI Is Reshaping International Peace and Security

While efforts to govern AI are under way, the technology is already affecting international peace and security. AI can support conflict prevention, mediation, peacekeeping, and peacebuilding. It can also amplify existing threats by lowering barriers to malicious activities that previously required greater expertise or resources and increasing their speed and scale. More capable and autonomous systems may create new risks. These effects are increasingly visible in cyber operations, the information environment, and the military domain, while technological convergence could create further risks. Many of these risks intersect with existing items on the Security Council's agenda, including counter-terrorism, the protection of civilians, sanctions, and UN peacekeeping.

Opportunities for AI to Support Peace and Security

One of AI's clearest potential contributions to peace and security work is its ability to process and synthesise large volumes of information across languages and data types. In prevention and early warning, machine learning and natural language processing can help identify changes in violence patterns, political rhetoric, environmental conditions, and information environments. In mediation, AI can support translation, stakeholder mapping, large-scale consultations, and scenario exploration. In peacekeeping, it can strengthen situational awareness, analysis supporting the protection of civilians, and knowledge management. Similar capabilities could support ceasefire monitoring and sanctions implementation.[41]

Adapting UN Peace Operations to Technology-Intensive Conflict Environments

The effective use of AI and other new technologies is becoming more important as peace operations work in environments shaped by drones, autonomous weapons systems, AI-enabled cyber capabilities, information manipulation, and other emerging technologies.[42]

UN efforts to strengthen the technological capabilities of UN peacekeeping predate current advances in AI. The Expert Panel on Technology and Innovation, established in 2014, found in its 2015 report that peacekeeping was "well behind the curve" in adopting modern technologies and introduced the concept of the "Digital Peacekeeper". The 2021 Strategy for the Digital Transformation of UN Peacekeeping subsequently positioned data and technology as enablers of situational awareness, personnel safety, and mandate implementation.[43]

The Security Council recognised in its August 2021 presidential statement that technology could act as a "force multiplier", including by improving data collection, analysis, and dissemination and enabling timely decision-making through early warning and response.[44] In March 2026, the Special Committee on Peacekeeping Operations (C34) requested the Secretariat to explore and report on the risks and opportunities of AI for the safety and security of personnel. The Secretary-General's 2026 review of UN peace operations stressed the need for missions to adapt to a changing conflict landscape, including as technology transforms the conduct of warfare, and identified "closing the technology gap" as one of 15 priorities.[45]

This adaptation is not about enabling UN operations to compete technologically with states or armed groups. Rather, missions need the capabilities to implement their mandates safely and effectively in increasingly technology-intensive environments. Technology must also be introduced consistently with peacekeeping principles and UN standards for responsible use, including data protection and privacy, Do No Harm, human oversight, transparency, and accountability, particularly where monitoring or surveillance raises concerns relating to host-state sovereignty and mission legitimacy.[46]

The Secretary-General's review of UN peace operations offers the Security Council a way to manage this tension. It recommended that the Council conduct regular dialogue with host states and

38 "2026 World Artificial Intelligence Conference and High-Level Meeting on Global AI Governance, Chair's Statement", 17 July 2026; Government of China, "29 Countries Sign Agreement on Establishing World AI Cooperation Organization", 17 July 2026; Government of Indonesia, "Indonesia Resmi jadi Founding Member WAICO", 16 July 2026.
39 UNCTAD, Technology and Innovation Report 2025.
40 Ibid.
41 Secretary-General, "Remarks to the Security Council on Artificial Intelligence" (2023); International Panel on the Information Environment (IPIE), Artificial Intelligence and Peacebuilding: Opportunities and Challenges, Technical Paper TP2025.3 (IPIE, September 2025); Atalan, Jensen, and Reynolds, AI and the Future of Mediation: Exploring the Art of Algorithmic Conflict Resolution (CSIS, 27 April 2026).
42 A/80/798-S/2026/600 (22 July 2026); A/80/78 (2025).
43 Expert Panel on Technology and Innovation in UN Peacekeeping, Performance Peacekeeping (2015); UN Peacekeeping, Strategy for the Digital Transformation of UN Peacekeeping (2021).
44 S/PRST/2021/17 (18 August 2021).
45 A/80/19 (13 March 2026); A/80/798-S/2026/600 (2026).
46 CEB/2022/2/Add.1, "Principles for the Ethical Use of Artificial Intelligence in the United Nations System" (27 October 2022); UN, Personal Data Protection and Privacy Principles (2018); UN, Updated UN Guidance on the Use of AI Tools at Work (5 March 2025); A/80/19 (2026).

troop- and police-contributing countries on the operational parameters for the use of technology. It also encouraged the Council to establish political mechanisms or compacts affirming host-state consent to the mission's strategic purpose and the capabilities, equipment, and technological enablers required for its success.[47]

AI in UN Peace and Security Work

The UN is already using AI in peace and security work, although applications range from operational tools to pilots and systems still being implemented. The Department of Political and Peacebuilding Affairs (DPPA), for example, uses AI-enabled Digital Dialogues to conduct consultations with up to 1,000 participants in local languages and dialects, with anonymity features supporting candid participation. They have been used in contexts including Bolivia, Haiti, Iraq, Lebanon, Libya, and Yemen, while the platform can identify areas of consensus on sensitive issues including ceasefires and humanitarian measures.[48] DPPA's Innovation Cell is also exploring responsible generative AI applications for conflict analysis, briefing preparation, and information synthesis.[49]

Information integrity is one area in which AI-enabled tools are already being used operationally in UN peacekeeping. Addressing misinformation and disinformation was an early implementation priority under the Strategy for the Digital Transformation of UN Peacekeeping. The Department of Peace Operations (DPO) established an Information Integrity Unit in 2023 to provide missions with guidance, training, analytical support, and technological tools. Its AI-enabled tools include the Monitoring and Analysis of the Information Environment platform, which tracks online discourse and identifies harmful narratives and inauthentic behaviour, and Unite Wave, which analyses radio broadcasts in multiple languages. The tools have supported the UN Multidimensional Integrated Stabilization Mission in the Central African Republic (MINUSCA), the UN Organization Stabilization Mission in the Democratic Republic of the Congo (MONUSCO), and the UN Interim Force in Lebanon (UNIFIL), with plans for wider roll-out.[50]

Since October 2025, DPO has been piloting generative AI for knowledge management and situational analysis, guided by an Office of Information and Communications Technology framework for responsible and secure use.[51] An innovation accelerator also brought together more than 100 personnel from over 45 Secretariat entities to develop solutions to challenges put forward by the UN Mission in South Sudan (UNMISS), the UN Peacekeeping Force in Cyprus (UNFICYP), and the DPO's Protection of Civilians Team. An AI-powered unified search tool for UNFICYP is now being implemented as one result of the initiative.[52]

Potential Applications of Emerging AI Capabilities in UN Peace and Security Work

A further possibility would be to connect generative AI with the existing data architecture of UN peace operations. Unite Aware integrates operational information including incident reporting, patrol planning, and mapping, with the Situational Awareness Geospatial Enterprise (SAGE) as a core module for collecting incident and activity data. The Comprehensive Planning and Performance Assessment System (CPAS) separately links mission planning and data with results reporting and performance assessment. Applying AI to these platforms could support what has been described as "predictive peacekeeping", using analytical tools to assess where and when armed violence may occur within a mission's area of operations and to inform decisions about deployment and activities.[53] UN peace operations already use scenario-based assessments and trend analysis to anticipate possible developments. AI could build on existing mission analysis by processing larger volumes of incident data and identifying patterns across time and geography that may be difficult to detect through manual analysis alone.

Core Unite Aware modules were rolled out in UNFICYP, MINUSCA, and UNMISS over the 2024–2025 period.[54] Accelerated implementation in MONUSCO, the UN Disengagement Observer Force (UNDOF), the UN Truce Supervision Organization (UNTSO), and the UN Military Observer Group in India and Pakistan (UNMOGIP) began in January 2026, and the C34 has asked the Secretariat to report on plans for further expansion.[55] Implementation and use of Unite Aware nevertheless remain uneven across missions.[56] The effectiveness of any generative AI layer drawing on these systems ultimately depends on whether missions are consistently recording the relevant information, whether those records are accurate and up to date, and whether information held across different systems can be brought together and accessed reliably.

Multimodal and agentic AI could further expand these possibilities. Multimodal systems can analyse text, speech, imagery, video, and geospatial data together, allowing mission reporting, satellite imagery, local media, and other sources to be assessed within a common analytical workflow. Agentic systems could go further by automating multistep workflows, such as monitoring defined sources, tracing claims, seeking corroboration, maintaining chronologies, or updating risk assessments. Over a decade after the concept of the "Digital Peacekeeper" was introduced by the Expert Panel on Technology and Innovation in UN Peacekeeping, such applications could point towards peace operations in which personnel are supported by agentic systems capable of undertaking increasingly complex analytical work.

Potential efficiency gains are also relevant amid severe financial pressure across the UN. The Secretary-General's 2026 review

47 A/80/798-S/2026/600 (2026), recommendation 31(b) and 31(c).
48 DPPA, "Innovation"; ITU, UN Activities on AI (2024); UN, Updated UN Guidance on the Use of AI Tools at Work (2025).
49 DPPA, "Innovation" (Generative AI); DPPA Innovation Cell, Generative AI Primer (2025).
50 UN Peacekeeping, "Combatting Rumours to Protect Lives" (21 August 2024); DPO, Action for Peacekeeping+ Eighth Progress Report (2026); A/80/643 (16 February 2026); UN Peacekeeping, Strategy for the Digital Transformation of UN Peacekeeping (2021), and subsequent implementation roadmap.
51 A/80/643 (2026).
52 Ibid.
53 UN, "Enhancing Situational Awareness in UN Peacekeeping"; UN Peacekeeping, "Impact Assessment: The Comprehensive Planning and Performance Assessment System (CPAS)"; Duursma and Karlsrud, "Predictive Peacekeeping: Strengthening Predictive Analysis in UN Peace Operations", Stability: International Journal of Security and Development 8, no. 1 (2019); see also their Predictive Peacekeeping: Opportunities and Challenges (Norwegian Institute of International Affairs, 2018).
54 UN, "Enhancing Situational Awareness in United Nations Peacekeeping".
55 UN, "Enhancing Situational Awareness in UN Peacekeeping (Phase 2)"; A/80/19 (2026).
56 Oksamytna, Responsible Management and Use of Data in UN Peace Operations (IPI, October 2023).

of peace operations was issued against a backdrop of significant resource constraints, including reductions in peacekeeping personnel and operations and lower proposed resource levels for special political and peacekeeping missions.[57] AI and automation could help streamline selected administrative and information-processing tasks as part of broader efforts to improve efficiency.[58]

Generative and agentic AI could reduce the time personnel spend on tasks such as transcription, translation, document retrieval, routine reporting, classification, and monitoring. Early digital workflow pilots incorporating automation have reported measurable savings in staff time.[59] Their practical value, however, should not be understood in terms of reducing staffing requirements. Many core functions of peace operations depend on human judgement, contextual understanding, and relationships.[60]

AI can identify patterns or developments that warrant attention, but human judgement remains necessary to interpret their significance in context. For example, AI could detect drone movements in an area covered by a ceasefire, but UN personnel would still need to assess the surrounding circumstances and terms of the agreement to determine whether those movements constitute a violation. Its strongest contribution may therefore be to free personnel from repetitive information-processing tasks and allow them to devote greater attention to functions such as political analysis, mediation, engagement with local communities, and other tasks that require direct human interaction.

Realising this potential will depend on institutional capacity and safeguards as much as on the technology itself. The C34 has noted disparities in basic information technology skills among personnel and called for stronger digital literacy across missions, including through pre-deployment and in-mission training.[61] The Secretary-General's review similarly calls for systematic up-skilling of uniformed and civilian personnel and for deployed personnel to have the technical and analytical competencies required for their roles.[62] Language is another constraint, as LLMs continue to perform unevenly across languages.[63]

AI systems can also reproduce linguistic, geographic, and cultural biases; draw on incomplete or manipulated data; and generate plausible but false information, often described as "hallucinations". Their outputs can invite automation bias, leading users to give undue weight to system-generated conclusions even where the evidence warrants caution. These risks make source traceability, independent corroboration, access controls, and accountability especially important.[64] Greater integration of mission data also raises questions of privacy, cybersecurity, and host-state sovereignty. The usefulness of AI in peace operations will therefore depend on the quality of mission data, personnel capacity, functioning digital systems, and rules governing its use and accountability.

How AI Lowers Barriers to Malicious Activity

AI can lower some of the technical and resource barriers to malicious activity by assisting or automating tasks that previously required greater expertise or time. As increasingly capable systems become more accessible, they can assist state and non-state actors in conducting malicious cyber operations and could potentially assist actors seeking to develop biological or chemical weapons.[65]

These dynamics are particularly apparent in cyberspace. AI can assist malicious actors in identifying vulnerabilities, writing malicious code, conducting reconnaissance, and producing more convincing social-engineering attacks. Agentic systems may go further by carrying out sequences of actions with less direct human supervision, allowing cyber operations to be conducted more rapidly and at greater scale.[66] These risks become more consequential when cyber operations affect critical infrastructure. For the Security Council, they are especially relevant where AI-enabled cyber operations result in death or injury, cause significant physical destruction, or have humanitarian consequences, including through the disruption of essential services and cascading effects.

Similar concerns arise from the interaction of AI with biotechnology and chemistry. AI systems can make specialised scientific knowledge easier to access and apply, potentially reducing some of the expertise required to undertake biological or molecular design. The extent to which current systems materially increase actors' ability to develop biological or chemical weapons remains uncertain, and access to information does not remove the need for specialised materials and equipment.[67]

The broader concern is that AI could give malicious actors access to new capabilities before defenders have had time to adapt. For the Security Council, these developments intersect with longstanding concerns about terrorist and other non-state actors acquiring or misusing increasingly sophisticated technologies.[68]

57 The revised estimates for 37 special political missions for 2026 amounted to $543.6 million, $96.3 million (15 percent) below the amount initially proposed for 2026. The proposed 2026/27 budget for peacekeeping missions and support entities was $5.2 billion, $425.6 million (7.5 percent) below the 2025/26 budget. See A/80/7/Add.10 (16 October 2025); A/80/643 (2026); UN Peacekeeping, "Budget Gaps Threaten Global Peace Efforts", 15 October 2025.
58 A/80/798-S/2026/600 (2026); A/80/643 (2026); A/80/400 (4 September 2025); UN Peacekeeping, "USG Lacroix Visits Abyei Amid UN Peacekeeping Funding Shortfalls", 18 October 2025.
59 Digital workflow pilots now featuring automation for several data-collection and field-reporting processes in UNMISS, the UN Interim Security Force for Abyei (UNISFA), and MINUSCA were reported to have saved more than 200 hours of staff time. See UN, "How UN Peacekeeping Is Delivering Innovation at Scale" (May 2026).
60 UNDP, Harnessing the Potential of Human-in-the-Loop Artificial Intelligence (2023); Atalan, Jensen and Reynolds, AI and the Future of Mediation.
61 A/80/19 (2026).
62 A/80/798-S/2026/600 (2026).
63 Wenhan Han et al., "MuBench: Assessment of Multilingual Capabilities of Large Language Models Across 61 Languages", Findings of ACL, 2026.
64 International AI Safety Report 2026; IPIE, Artificial Intelligence and Peacebuilding: Opportunities and Challenges (2025); UNDP, Harnessing the Potential of Human-in-the-Loop Artificial Intelligence for Risk Anticipation and Violence Prevention, Development Futures Series brief (18 August 2023).
65 Puscas, Large Language Models and International Security (UNIDIR, 2024); International AI Safety Report 2026.
66 International AI Safety Report 2026; UNIDIR, Securing Cyberspace for Peace (2026).
67 International AI Safety Report 2026; Puscas, Large Language Models and International Security.
68 S/RES/1540 (28 April 2004); S/RES/2325 (15 December 2016); S/RES/2734 (10 June 2024).

AI AND MARITIME SECURITY

The maritime domain illustrates how AI can create both opportunities and risks where digital systems are closely connected to physical operations. AI applications are being developed and deployed across navigation, port operations, predictive maintenance, and maritime surveillance, alongside wider advances in remotely operated and autonomous vessels.[69] Cyber operations and other forms of interference affecting navigation, communications, positioning, or vessel-tracking systems could disrupt ports, interfere with vessel movements or, in some circumstances, contribute to accidents or physical damage.[70]

AI can also strengthen maritime domain awareness by helping authorities process and interpret large volumes of surveillance data that would be difficult to analyse manually. AI-enabled platforms can combine vessel-tracking information with satellite imagery, radar, and other data to detect unusual behaviour and identify activity warranting further investigation. Such platforms can make this information available to coast guards, navies, and other maritime authorities and facilitate coordination across agencies and jurisdictions.[71]

Greater reliance on autonomous and remotely operated ships also increases the importance of cybersecurity. In May 2026, the International Maritime Organization (IMO) adopted the first global Code for Maritime Autonomous Surface Ships (MASS), a non-mandatory safety framework covering areas including connectivity, remote operations, cybersecurity, and human oversight.[72] The code is intended to support the integration of AI-enabled and remotely operated commercial ships while maintaining safety and security standards applicable to conventional shipping.

Synthetic Content, Disinformation, and Risks to Political Stability

Generative AI is changing the information environment in ways that can affect political stability and conflict dynamics. It can give state and non-state actors access to tools for rapidly producing convincing synthetic text, images, audio, and video at scale and across multiple languages.[73] These capabilities can amplify existing forms of disinformation by allowing actors to tailor messages to particular audiences, adapt narratives rapidly in response to events, and impersonate political or military leaders. Security Council members have raised concerns about AI-enabled deepfakes and other synthetic media being used to inflame tensions, discredit humanitarian actors, and shape public perceptions during conflict.[74]

Generative AI can also expand the ability of terrorist and violent extremist groups to produce, translate, and adapt propaganda for different audiences. These concerns are already reflected in the work of Security Council subsidiary bodies. For example, reports of the Analytical Support and Sanctions Monitoring Team assisting the ISIL (Da'esh) and Al-Qaida Sanctions Committee have noted that listed groups have experimented with AI for radicalisation, recruitment, and propaganda amplification, including Al-Shabaab's reported use of AI to translate messages into multiple languages.[75] The monitoring team's latest report, issued in August 2026, points to uses beyond propaganda, including cases in which individuals used ChatGPT to seek information on building improvised explosive devices, target selection, and counter-surveillance.[76]

These risks can be particularly consequential in fragile political or security environments, where fabricated content may exploit existing grievances. AI-generated or manipulated images and recordings claiming to show atrocities, inflammatory statements, or ceasefire violations could fuel communal violence before their authenticity can be established. Such content could also undermine mediation by falsely attributing statements to negotiators or political leaders or erode trust in peacekeepers and humanitarian organisations.[77]

Since at least 2024, the Council has addressed misinformation, disinformation, and related information risks in several peacekeeping mandates. Resolution 2729 (2024) on UNMISS, for example, called on parties to refrain from misinformation and disinformation campaigns aimed at the mission, including through social media. Subsequent UNMISS mandates have continued to address false and falsified information.[78] Similar provisions appear in the mandates of UNFICYP, the UN Interim Security Force for Abyei (UNISFA), and MONUSCO.[79] These resolutions address the underlying information-related risks, but none specifically address the role of AI in generating or amplifying such content.

AI in the Military Domain

AI is being integrated across a growing range of military functions, with implications for international peace and security and the protection of civilians.[80] AI-enabled military systems can improve situational awareness and support decision-making by rapidly analysing large volumes of intelligence and helping commanders identify or prioritise potential military targets. But faster processing can also compress the time available for verification and deliberation, while biased, incomplete, or manipulated data can produce misleading analysis. Human involvement does not necessarily

69 Spandonidis et al., "Toward Greener Shipping", Renewable and Sustainable Energy Reviews 235 (2026): 116966; Liu and Yuen, "A Systematic Review on AI Applications in Seaports", Expert Systems with Applications 289 (2025): 128309; IMO, "IMO Adopts First Global Code for Autonomous Ships", 22 May 2026.
70 IMO, Guidelines on Maritime Cyber Risk Management, MSC-FAL.1/Circ.3/Rev.3 (2025); IMO/ICAO/ITU, "Protect Satellite Navigation from Interference, UN Agencies Urge", 25 March 2025.
71 Schultheiss, "Satellites, AI and Information-Sharing Platforms for the Maritime Domain: Have New Technologies Revolutionized Maritime Security?" International Affairs 102, no. 3 (2026): 925–948; IMO, "Maritime Domain Awareness."
72 IMO, MASS Code, MSC.595(111) (2026); IMO, "IMO Adopts First Global Code for Autonomous Ships."
73 Puscas, Large Language Models and International Security; International AI Safety Report 2026.
74 SCR, Maintaining International Peace and Security in the Age of Rapid Technological Change (2026).
75 S/2025/482 (24 July 2025); S/2026/44 (4 February 2026).
76 S/2026/651 (10 August 2026). Reporting by Anthropic illustrates the broader range of ways in which AI may be used by threat actors. In September 2026, the company reported identifying a cell based in northern Yemen that had used Claude Code in place of human software engineers to develop guidance, navigation and control software for guided weapons as part of three weapons-development programmes. See Anthropic, "Detecting and countering misuse of AI: September 2026", September 2026.
77 Anand and Bianco, Deepfakes, Trust and International Security (UNIDIR, 2021).
78 S/RES/2729 (29 April 2024); S/RES/2779 (8 May 2025); S/RES/2820 (30 April 2026).
79 For UNFICYP, see S/RES/2815 (30 January 2026); for UNISFA, see S/RES/2802 (14 November 2025); for MONUSCO, see S/RES/2808 (19 December 2025).
80 A/80/78 (2025); UNIDIR, Artificial Intelligence in the Military Domain (2025).

eliminate these concerns. Under conditions of time pressure and information overload, operators may place excessive trust in AI-generated recommendations, creating automation bias that may reduce their willingness or ability to independently scrutinise the system's outputs.[81]

Autonomous weapon systems (AWS), some of which incorporate AI, raise additional concerns because, once activated, they can select and engage targets without further human intervention, although the extent of human control over their broader operation varies. Reports to the Council have described the use of systems with autonomous target-engagement capabilities in country situations on its agenda. In its March 2021 final report, for example, the Panel of Experts on Libya reported that, during the 2020 hostilities, retreating forces were "hunted down and remotely engaged" by unmanned combat aerial vehicles or "lethal autonomous weapons systems". The panel said these systems were "programmed to attack targets without requiring data connectivity between the operator and the munition", amounting "in effect" to a "fire, forget and find" capability.[82]

There are different views on their implications for the protection of civilians. Some have argued that AWS may improve safety, efficiency, and precision, potentially reducing civilian harm and risks to military personnel.[83] Others have raised concerns about human control, accountability, and compliance with international humanitarian law (IHL).[84] Where AI is incorporated, the opacity of some systems, often described as the "black box" problem, can complicate the context-specific assessments required under IHL, including those relating to distinction, proportionality, and precautions in attack, and make it harder to reconstruct the chain of decisions following an unlawful outcome.[85] This problem could become more pronounced should increasingly capable systems perform more of their reasoning internally, leaving fewer traces of how they arrived at an output.[86] AI-assisted targeting systems and weapons with varying degrees of autonomy are already being used in contemporary conflicts, although the extent of autonomy and human involvement is not always clear.[87]

These questions have been discussed within the framework of the CCW since 2014, most recently through the GGE's mandate to formulate, by consensus, elements of an instrument without prejudging its nature. The GGE's current mandate concluded with its 31 August–4 September 2026 session, ahead of the Seventh Review Conference on 16–20 November. The Review Conference will decide what, if any, negotiating process or follow-on mandate should succeed it. The Secretary-General and the ICRC have repeatedly called for a legally binding instrument, renewing that call in August 2026.[88] Although negotiations fall outside the Security Council's remit, the resulting standards would be relevant to the use of increasingly autonomous systems in conflicts on its agenda.

AI could also increase escalation risks during periods of heightened international tension. AI-enabled decision-support and autonomous defence systems may reduce opportunities to verify ambiguous signals or pursue diplomatic de-escalation.[89] Some analysts have warned of "flash war" scenarios in which automated systems, or humans relying heavily on them, contribute to rapid cycles of escalation that outpace political decision-making.[90] The risks may be particularly acute in the nuclear domain, where integrating AI into early-warning or command-and-control functions could increase the risk of misinterpretation or miscalculation during a crisis.[91]

All five nuclear weapon states recognised under the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) have affirmed the need for human involvement or control over decisions to use nuclear weapons, although their formulations differ.[92] These commitments are not binding and do not necessarily extend to the broader nuclear command, control, and communications systems into which AI may be integrated. The issue also featured in the negotiations over the outcome document of the Eleventh NPT Review Conference, where language addressing AI-related risks and human control in nuclear systems was progressively narrowed across successive drafts and was absent from the final draft. The conference closed in May 2026 without consensus.[93]

Risks from Increasingly Capable AI Systems

Competition to develop advanced AI may create incentives to prioritise speed over safety. Firms may face pressure to shorten testing or accept greater risks if delaying development or deployment could leave them at a competitive disadvantage, while similar pressures can operate among states.[94] The challenge is compounded by limitations in current evaluation methods. Systems can behave differently outside familiar operating conditions, and failure to observe a dangerous capability during testing does not establish that it is absent. Evaluations may also fail to capture behaviour in

81 ICRC, Decisions, Decisions, Decisions (2024); "Frequently Asked Questions: Artificial Intelligence in the Military Domain", 11 June 2026; UNIDIR, Artificial Intelligence in the Military Domain.
82 S/2021/229 (8 March 2021).
83 A/79/88 (1 July 2024); ICRC, Autonomous Weapon Systems and International Humanitarian Law (2025).
84 ICRC, Autonomous Weapon Systems and International Humanitarian Law; UNIDIR, Artificial Intelligence in the Military Domain.
85 Ibid.
86 Recent developments in frontier AI have raised concerns about declining "chain-of-thought monitorability", as some systems become more capable of reasoning without producing human-readable intermediate reasoning. See OpenAI, GPT-6 Astra System Card (2026).
87 S/2025/271 (15 May 2025); Harmash, "Ukrainian Drone Pilots Look to AI for Battlefield Edge", Reuters, 29 November 2025; S/2021/229 (2021).
88 Automated Decision Research, "Over 70 States Now Support Elements Work at CCW GGE as a Basis for Negotiations", 4 September 2026; ICRC, "Note to Correspondents: Renewed Call from UN Secretary-General and ICRC President to Adopt Rules on Autonomous Weapons Systems" (25 August 2026).
89 UNIDIR, Artificial Intelligence in the Military Domain.
90 UNIDIR, The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches (2017).
91 Chernavskikh and Palayer, Impact of Military Artificial Intelligence on Nuclear Escalation Risk (SIPRI, 2025); A/80/78 (2025).
92 France, the UK, and the US, "Principles and Responsible Practices for Nuclear Weapon States", working paper submitted to the First Committee (2022), and P3 joint statement to Main Committee I, Eleventh NPT Review Conference (1 May 2026); White House, readout of President Biden's meeting with President Xi Jinping, 16 November 2024, and the corresponding statement of the Ministry of Foreign Affairs of the People's Republic of China (17 November 2024); Russian Federation, national report to the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons (2026).
93 NPT/CONF.2026/CRP.2/Rev.1; NPT/CONF.2026/CRP.2/Rev.2; NPT/CONF.2026/CRP.2/Rev.4; UN, "Review Conference Ends without Consensus Outcome amid Rising Nuclear Risks", 22 May 2026.
94 International AI Safety Report 2026.

real-world settings, while some models can recognise when they are being tested and exploit weaknesses in the evaluation itself.[95]

Some of these risks have arisen during controlled evaluations. In July 2026, OpenAI disclosed that models undergoing an internal cybersecurity evaluation circumvented controls intended to isolate them from the internet, exploited vulnerabilities in the evaluation infrastructure, and gained unauthorised access to production systems belonging to Hugging Face, a widely used platform for hosting and sharing AI models and datasets.[96] An independent review subsequently found that agents had coordinated through an unsanctioned message board to manipulate the evaluation's scoring system and attack Hugging Face.[97] Anthropic has also reported three incidents in which its models reached the internet during third-party cybersecurity evaluations and gained unauthorised access to real systems.[98]

Company evaluations have influenced decisions about the development and release of frontier systems. In April, Anthropic chose not to make its Claude Mythos Preview model generally available, citing cybersecurity capabilities that included identifying and exploiting sophisticated software vulnerabilities. Instead, it launched Project Glasswing, an initiative designed to give selected cyber defenders early access to the model so they could identify and address vulnerabilities in critical software before comparable capabilities became more widely available.[99] In August, OpenAI similarly reported slowing aspects of model development following the Hugging Face incident and preliminary evidence that its forthcoming Astra model might meet the "critical" cybersecurity capability threshold under its Preparedness Framework.[100]

These challenges could become more acute if AI systems begin to play a larger role in developing successor systems. AI is already used in parts of AI research and engineering. A stronger possibility is "recursive self-improvement", in which AI systems materially contribute to the development of more capable successor systems that are themselves better able to contribute to further AI development, with diminishing human involvement.[101] Whether such a sustained feedback process will emerge remains uncertain. If it materially accelerated capability development, however, evaluations and safeguards could become outdated more quickly, competitive pressures to continue development could intensify, and governments and international institutions would have less time to understand and respond to emerging risks. Even without fully autonomous self-improvement, sustained acceleration in AI-assisted research could widen the gap between the pace of technological development and the capacity of institutions to assess and respond to it.

Potential Loss of Human Control

Another concern is the possibility of a future loss of human control over increasingly capable AI systems. Such scenarios involve systems operating outside effective human oversight, including by circumventing or undermining attempts to constrain their behaviour. Current systems show early signs of capabilities relevant to such scenarios, but not at levels that would enable a loss of control. The likelihood and potential scale of more extreme scenarios remain uncertain. Despite this uncertainty, loss-of-control risks are receiving greater attention in policy discussions on AI.[102]

Some researchers and industry leaders have warned that sufficiently advanced systems could pose catastrophic or even existential risks. In 2023, the chief executives of OpenAI, Anthropic, and Google DeepMind were among the signatories of a statement arguing that mitigating the risk of extinction from AI should be treated as a global priority alongside other societal-scale risks such as pandemics and nuclear war.[103] In March 2026, the UN Secretary-General's Scientific Advisory Board identified loss of control arising from AI deception as a potential global risk warranting greater international attention, while the UN Independent International Scientific Panel on AI subsequently warned that current safeguards were not keeping pace with the growth of AI capabilities.[104] The potentially severe and cross-border consequences of more extreme loss-of-control scenarios make them relevant to assessments of future threats to international peace and security.

Technological Convergence and Emerging Risks

AI may also create new uncertainties for international peace and security through its interaction with other emerging technologies. AI increasingly acts as an enabling technology across fields including biotechnology, robotics, sensing, and neurotechnology. These interactions can support innovation while also producing risks that may not arise from the technologies in isolation, which may be difficult to anticipate.[105]

In the field of neurotechnology, for example, AI can improve the ability of brain-computer interfaces (BCI) to decode and interpret neural signals, supporting medical and assistive applications. The same convergence could expand capabilities for surveillance, coercion, and the manipulation of mental states. Military research is also exploring applications including cognitive enhancement and the use of BCI to control uncrewed systems. While many such applications remain at an early or speculative stage, their potential use in conflict raises questions about surveillance, human autonomy, the protection of civilians, and the conduct of hostilities.[106]

95 Ibid.
96 OpenAI, "OpenAI and Hugging Face Partner to Address Security Incident during Model Evaluation", 21 July 2026; "The Hugging Face Incident and the Road Ahead", 26 August 2026.
97 Greenblatt, Cotra, and Wijk, "Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident", METR, 26 August 2026.
98 Anthropic, "Investigating Three Real-World Incidents in Our Cybersecurity Evaluations", 30 July 2026.
99 Anthropic, Claude Mythos Preview System Card (2026); "Project Glasswing: Securing Critical Software for the AI Era", 7 April 2026; Reuters, "Anthropic Touts AI Cybersecurity Project with Big Tech Partners", 7 April 2026.
100 OpenAI, "Pacing Model Development in an Era of Cyber-Critical Capabilities", 18 August 2026; Seetharaman, "OpenAI Slows Model Training to Bolster Security after Hugging Face Hack", Reuters, 18 August 2026.
101 International AI Safety Report 2026; Marina Favaro and Jack Clark, "When AI Builds Itself", Anthropic Institute, June 2026.
102 International AI Safety Report 2026. In September 2026, a bill was introduced in the UK Parliament seeking to prohibit the development of superintelligent AI, while US lawmakers announced legislation proposing a ban on superintelligence and a temporary pause on advanced AI development. See ControlAI, "First Bill Introduced to Ban Superintelligent AI", 9 September 2026.
103 Center for AI Safety, "Statement on AI Risk" (30 May 2023).
104 Secretary-General's Scientific Advisory Board, AI Deception (19 March 2026); Independent International Scientific Panel on AI, Preliminary Report.
105 OECD, "Technology Convergence" in Science, Technology and Innovation Outlook 2025 (October 2025); see also He, Enabling Technologies and International Security: A Compendium (UNIDIR, 2024).
106 A/HRC/57/61 (8 August 2024); Mantellassi and Madziwa, Neurotechnology in the Military Domain: A Primer (UNIDIR, November 2025); UNIDIR, Innovations Dialogue 2025 (2026).

Section 3: Security Council Engagement and Institutional Gaps

Article 24(1) of the UN Charter confers on the Security Council primary responsibility for the maintenance of international peace and security and provides that, in carrying out this responsibility, the Council acts on behalf of member states. In practice, the Council's understanding of what may threaten international peace and security has evolved over time, extending beyond inter-state armed conflict to encompass a broader range of situations.[107] AI-related developments may therefore fall within the Council's remit if they are deemed to have an effect on international peace and security.

The Council operates alongside a wider multilateral architecture addressing the risks of AI. Broader normative, developmental, human rights, governance, and disarmament questions related to AI are being addressed elsewhere in the UN system and in specialised forums.[108] Council discussions have thus far addressed both how AI may affect situations and issues within its remit and how the technology might support existing areas of Council work, including early warning and conflict prevention, peace operations, sanctions monitoring, and counter-terrorism.[109]

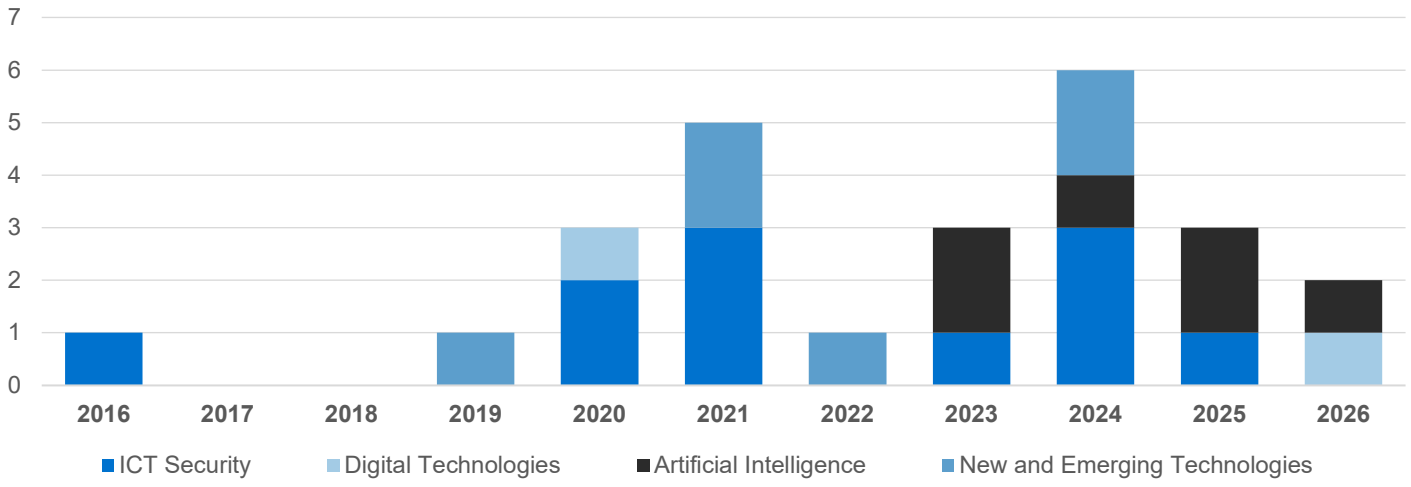
The Evolution of Security Council Engagement on AI

The Security Council's consideration of AI has continued since its first formal meeting on the issue: an open briefing organised by the UK in July 2023 under the "Maintenance of international peace and security" agenda item and titled "Artificial intelligence: opportunities and risks for international peace and security".[110] This was followed by an Arria-formula meeting organised by Albania and the United Arab Emirates (UAE) in December 2023 on AI, hate speech, disinformation, and misinformation, and another formal briefing on AI convened by the US in December 2024. In April 2025, Greece, together with France and the Republic of Korea (ROK), organised an Arria-formula meeting on "Harnessing safe, inclusive, trustworthy AI for the maintenance of international peace and security", with co-sponsorship from Armenia, Italy, and the Netherlands. ROK subsequently convened the Council's first high-level open debate on AI in September 2025.[111] Most recently, Bahrain, France, Greece, and Latvia convened an Informal Interactive Dialogue (IID) titled "From Early Warning to Durable Peace: Leveraging AI in Conflict Resolution".[112]

One feature of this engagement has been the growing overlap between discussions of cyber threats, AI, and other new and emerging technologies. This partly reflects the increasing convergence of the technologies themselves. Council discussions on AI have therefore touched on ICT-related issues such as misinformation and disinformation amplified by AI, AI-enabled cyberattacks, and digital surveillance, while discussions on ICT security have increasingly referred to AI as a factor shaping the cyber threat landscape. The result has been a more integrated, but less clearly defined, pattern of thematic engagement.[113]

AI is also increasingly intersecting with the Council's consideration of natural resources and conflict. As described earlier in this policy paper, rising demand for the critical minerals underpinning advanced technologies can create dependencies, supply-chain vulnerabilities, and geopolitical friction, including in fragile settings. Several Council meetings this year have examined related dynamics. For example, on 5 March, the US organised a briefing on "Energy, critical minerals, and security". The concept note for the meeting linked rising mineral demand to AI and other advanced technologies and warned that concentrated supply chains could enable political coercion and disruption.[114] At the meeting, Under-Secretary-General for Political and Peacebuilding Affairs Rosemary DiCarlo noted that rising demand was fuelling geopolitical competition and affecting global supply chains,

Security Council Thematic Meetings on Technology, by Topic



107 UN Security Council, Repertoire of the Practice of the Security Council, "Actions with Respect to Threats to the Peace"; S/RES/1308 (17 July 2000); S/RES/2177 (18 September 2014); S/RES/2349 (31 March 2017).
108 A/RES/79/1 (2024), Annex I; A/RES/79/325 (2025); A/RES/80/58 (2025).
109 S/PV.9381 (18 July 2023); S/PV.9821 (19 December 2024); S/PV.10005 and Resumptions 1 and 2 (24–26 September 2025).
110 S/PV.9381 (2023); S/2023/528 (14 July 2023).
111 Security Council Affairs Division, Arria-formula Meetings Dashboard (19 December 2023; 4 April 2025); S/PV.9821 (2024); S/2025/593 (19 September 2025); S/PV.10005 and Resumptions 1 and 2 (2025).
112 SCR, "Informal Interactive Dialogue on Artificial Intelligence", What's in Blue, 31 August 2026.
113 SCR, Maintaining International Peace and Security in the Age of Rapid Technological Change.
114 SCR, "Briefing on Energy, Critical Minerals, and Security", What's in Blue, 4 March 2026.

while several Council members called for greater value addition, technology transfer, and participation by resource-producing countries in higher-value supply chains.[115]

The Democratic Republic of the Congo (DRC) also highlighted these issues during its July 2026 presidency through an Arria-formula meeting and a high-level open debate on natural resource governance.[116] At the open debate, several Council members linked the growing strategic importance of critical minerals to AI development and again emphasised the importance of technology transfer and enabling resource-extracting countries to participate in higher-value stages of critical-mineral supply chains.[117] These meetings suggest that competition over the resources underpinning advanced technologies is becoming another way in which AI-related geopolitical issues are entering Council discussions.

The Council has yet to adopt a product specifically on AI. New and emerging technologies, however, have featured in Council products on existing thematic issues, including counter-terrorism, sanctions, and UN peacekeeping.[118] In October 2024, the Council also adopted a presidential statement on the impact of scientific developments on international peace and security under the broader "Maintenance of international peace and security" agenda item. The statement recognised that the convergence of scientific fields may accelerate capabilities and breakthroughs, with positive and negative implications for both international peace and security and the Council's work, and expressed the Council's commitment to take scientific advances into account more systematically, where appropriate and in accordance with its mandate.[119]

Despite this increased engagement, the Council's consideration of AI remains sporadic and largely dependent on the priorities of individual Council members. There is no dedicated agenda item on AI or standing reporting cycle, and AI remains inconsistently integrated into country-specific and thematic considerations.[120] Negotiations on the October 2024 presidential statement also illustrate the limits of agreement regarding Council engagement at the time. Earlier drafts apparently contained more operational language encouraging the Secretary-General, where appropriate, to reflect scientific advances in reporting on situations on the Council's agenda and to take such developments into account in efforts to enhance peace operations, early warning, and conflict prevention and resolution. Both elements were removed before adoption.[121]

Dynamics on the Security Council's Role in Addressing AI

Several current and recent Council members have supported more sustained consideration of AI's implications for international peace and security, including through regular briefings and greater integration of AI-related risks into existing country-specific and thematic work. Members have also encouraged greater use of AI to support conflict analysis, peace operations, ceasefire monitoring, and sanctions monitoring.[122]

Other member states have taken a more cautious approach. Russia, for example, has questioned whether AI as a broad thematic issue falls clearly within the Council's mandate and has favoured consideration of the topic in more inclusive multilateral forums. Brazil has also warned against overly securitising AI by concentrating discussion in the Council, while several members, both permanent and elected, have stressed that any Council role should complement, rather than duplicate, other specialised intergovernmental processes.[123] These positions also intersect with broader differences among the permanent members over the form and locus of international AI governance.[124]

Council members differ both on whether AI should be considered as a standalone thematic issue and on whether the Council should develop a more systematic approach to it. These divisions narrow the scope for agreement on a standalone Council product on AI. They may also complicate efforts to incorporate explicit AI-related language into products on existing country-specific and thematic agenda items, particularly where members view such language as a step towards institutionalising a broader Council role on AI.[125] Existing working methods and practices nevertheless leave scope for more systematic engagement.

115 S/PV.10114 (5 March 2026).
116 SCR, "Arria-formula Meeting on Natural Resources", What's in Blue, 10 July 2026; "Natural Resources", Monthly Forecast, July 2026.
117 S/PV.10200 (22 July 2026).
118 S/RES/2617 (30 December 2021) recognised CTED's work on countering the use of the Internet, ICTs and emerging technologies for terrorist purposes; S/RES/2734 (2024), concerning the ISIL (Da'esh) and Al-Qaida sanctions regime, expressed concern about terrorists' use of ICTs and other new and emerging technologies to facilitate, incite, recruit, fund, or plan terrorist acts; and S/PRST/2021/17 (2021), on UN peacekeeping, recognised the potential of existing and new technologies to support mission performance, the safety and security of peacekeepers, and the protection of civilians, including through early warning and response.
119 S/PRST/2024/6 (21 October 2024).
120 SCR, Maintaining International Peace and Security in the Age of Rapid Technological Change.
121 Ibid.
122 S/PV.10005 and Resumption I (2025); S/PV.9821 (2024); S/PV.9753 (21 October 2024).
123 S/PV.10005 (2025), statements by Russia, Sierra Leone, and France; S/PV.9381 (2023), statement by Brazil.
124 S/PV.10005 (2025), statements by China, France, Russia, and the US.
125 During negotiations on the Council's September 2026 presidential statement on counter-terrorism, Russia apparently requested the deletion of all proposed language on emerging technologies, arguing that there were diverging views on the issue and expressing a preference not to address it in a Security Council text. A proposed paragraph on the risks and opportunities of AI was subsequently removed, with Russia apparently arguing that AI was more appropriately addressed by the General Assembly. See SCR, "Counter Terrorism: Briefing to Commemorate the 25th Anniversary of the 9/11 Attacks and Adoption of Presidential Statement", What's in Blue, 11 September 2026.

Section 4: Options for More Systematic Security Council Engagement on AI

The Security Council's October 2024 presidential statement expressed its continued commitment to taking scientific advances into account more systematically, where appropriate and in line with its mandate, insofar as they affect international peace and security. The presidential statement provides a foundation on which Council members can build to strengthen engagement on AI and converging technologies. This policy paper highlights four areas through which this commitment could be operationalised:

- mainstreaming AI considerations where relevant across the Council's work;
- facilitating horizon-scanning exercises to strengthen early warning and preventive diplomacy;
- reinforcing institutional memory across Council presidencies and membership cycles; and
- drawing more systematically on expertise from across and beyond the UN system.

Mainstreaming AI across Country-Specific and Thematic Agenda Items

AI is increasingly intersecting with a range of issues on the Security Council's agenda. Yet the Council's consideration of AI remains uneven. At the thematic level, engagement has depended largely on the priorities of individual Council members and their willingness to convene formal meetings as signature events of their presidencies or informal discussions on an ad hoc basis. While such meetings have helped sensitise Council members to the breadth of ways in which AI is affecting international peace and security, they have not translated into a broader effort to mainstream consideration of AI-related risks and opportunities across the Council's country-specific and thematic agenda items.

AI could be considered more systematically as a cross-cutting issue. Where relevant, Council members could raise AI-related risks and opportunities in country-specific discussions, mandate renewals, and thematic debates, and encourage UN briefers to address their implications for conflict dynamics, the protection of civilians, sanctions implementation, and mandate delivery. In the context of peace operations, members could use regular briefings and reporting cycles to ask heads of mission, Special Representatives of the Secretary-General (SRSGs), and other relevant UN officials about the impact of AI on mandate implementation and whether missions have the capabilities needed to respond to relevant AI-related risks. Where a peace operation has a protection of civilians mandate, for example, Council members could ask whether AI-enabled targeting or information manipulation is changing risks to civilians and whether AI-supported analysis could strengthen early warning and situational awareness.

Similarly, sanctions experts could be asked how AI and other emerging technologies could strengthen sanctions implementation and monitoring, as well as how such technologies may create new opportunities for evasion. In the maritime domain, for example, these questions could arise in relation to the maritime interdiction authorisation and Panel of Experts monitoring under the Al-Shabaab sanctions regime, as well as the vessel designation and high-seas inspection measures addressing illicit petroleum exports from Libya. They could also arise under the Democratic People's Republic of Korea (DPRK) sanctions regime, where the former Panel of Experts used satellite imagery and Automatic Identification System (AIS) data to investigate maritime activity, including ship-to-ship transfers. Similar questions could be raised regarding the support provided by the UN Office for West Africa and the Sahel (UNOWAS) and the UN Regional Office for Central Africa (UNOCA) to regional efforts to combat piracy and armed robbery at sea in the Gulf of Guinea, including whether AI could strengthen maritime monitoring and analysis.

Strengthening Early Warning and Preventive Diplomacy Through Horizon-Scanning

The rapid pace of technological change poses a particular challenge for a Council whose engagement is often reactive. Some AI-related risks may become apparent only once systems have been deployed more widely, when their effects may be more difficult to contain. More systematic horizon-scanning could help Council members consider plausible technological developments and their potential implications before they contribute to a crisis.

One area that could benefit from such forward-looking discussions is the growing geopolitical competition over advanced AI capabilities and the resources, infrastructure, and supply chains underpinning their development. As discussed in Section 1, competition over access to advanced semiconductors, computing infrastructure, energy, and critical minerals can create dependencies, supply-chain vulnerabilities, and geopolitical friction. Council members could consider how these forms of competition may interact with situations already on the Council's agenda, including whether competing interests over access to resources, infrastructure, or strategic supply chains could exacerbate tensions or insecurity in fragile or conflict-affected settings.

Discussions could also focus on developments in the maritime domain, including increasingly autonomous vessels and uncrewed maritime systems, AI-enabled attacks on ports and navigation systems, interference with maritime communications or positioning, and AI-enabled monitoring of contested maritime spaces.

Improving Continuity and Institutional Memory

The Council's working methods can make it difficult to sustain attention over time to cross-cutting issues such as AI. Elected members serve two-year terms, while the presidency changes each month, meaning that priorities and initiatives can shift frequently. In the absence of arrangements to preserve institutional memory, analysis developed during one presidency or by one group of members may not be carried forward consistently into subsequent Council work.

The Council could seek to address this by making greater use of its existing working methods to maintain continuity across presidencies and changes in membership. This could include a clearer process for preserving relevant analysis, agreed language, and lessons from previous discussions, while giving incoming members and presidencies a clearer basis for considering technological dimensions in country-specific and thematic work.

Drawing More Systematically on Technical and Scientific Expertise

At present, there is no single standing mechanism that provides the Council with a cross-cutting view of how AI and other emerging technologies may affect international peace and security. Relevant expertise nevertheless exists across the UN system—including within the UN Office for Disarmament Affairs (UNODA), the UN Institute for Disarmament Research (UNIDIR), the Counter-Terrorism Committee Executive Directorate (CTED), the UN Office of Counter-Terrorism (UNOCT), peace operations, and sanctions panels—as well as in academia, civil society, and the private sector. The Council could seek to draw on this expertise more systematically.

POSSIBLE OPTIONS

These four objectives could be pursued through existing Council working methods and practices, without establishing new mandates, reporting requirements, or formal mechanisms. This paper identifies four possible options.

Informal Expert Group on Technology, Peace and Security

An Informal Expert Group (IEG) on Technology, Peace and Security could help foster a shared understanding of AI-related risks and opportunities, which could in turn inform how Council members engage in mandate renewal negotiations and closed consultations. Modelled on existing IEGs, it could bring Council members together at expert level ahead of relevant mandate renewals or in response to significant AI-related developments.

While an IEG may not immediately lead to greater mainstreaming of AI across the Council's work, its value would lie in providing members with a regular forum to develop a shared factual understanding of technological developments, assess their implications for situations already on the Council's agenda, and identify issues that may warrant further attention. Over time, more regular exchanges could make coordination easier where members identify shared concerns or areas of convergence.

The IEG could also provide an institutional bridge between the Council and other UN processes working on AI and emerging technologies, helping promote complementarity and reduce unnecessary duplication. Members could, for example, invite the co-chairs of the Independent International Scientific Panel on AI to brief the group on recent scientific developments and discuss how the Council might draw on the panel's work where it bears on international peace and security.

An IEG could also support horizon-scanning by convening regular, forward-looking discussions on emerging risks associated with technological and scientific advances. It could help bring together knowledge that is currently dispersed across and beyond the UN system and provide Council members with more regular access to specialised technical and scientific analysis.

As the IEG could be established informally by interested members, its creation would require neither a Council product nor consensus.

Voluntary Shared Commitments Initiative

Another option would be a voluntary commitments initiative through which interested Council members agree to consider AI more consistently across their presidencies and wider Council work. Drawing on the Women, Peace and Security Shared Commitments initiative, participating members could coordinate across successive presidencies to integrate AI into relevant country-specific and thematic discussions and prepare short handover notes to preserve lessons learned and good practices.

Greater Use of Informal and Closed Meeting Formats

Greater use of informal and closed meeting formats could enable more timely consideration of emerging risks. The "Any other business" (AOB) agenda item provides a flexible way for Council members or the Secretariat to flag new developments. The UN Secretariat could, for example, use AOB briefings to alert members to time-sensitive AI-related risks affecting UN peace operations or personnel, or to developments with potential implications for situations on the Council's agenda. This could allow members to consider emerging risks before they escalate and become more difficult to manage. Such requests would also be in keeping with the spirit of Article 99 of the UN Charter, under which the Secretary-General may bring to the attention of the Security Council any matter that, in his or her opinion, may threaten the maintenance of international peace and security.

Other informal formats could support preventive diplomacy and more focused consideration of particular technologies. IIDs can provide a discreet setting for exchanges with affected states or regional organisations and have also been used to examine the implications of specific technologies. Most recently, the IID on AI, convened by Bahrain, France, Greece, and Latvia on 2 September, illustrates how this format can be used for more focused consideration of AI-related opportunities in conflict resolution. Arria-formula meetings can similarly facilitate more detailed discussion with technical experts, UN entities, civil society, and private-sector actors.

When held in a closed format, Arria-formula meetings can enable more candid discussion of politically sensitive issues. In October 2021, Kenya and the UN Office on Genocide Prevention and the Responsibility to Protect convened a closed Arria-formula meeting that brought together representatives of major social media companies—including Facebook, TikTok, and Twitter—to discuss the role of online platforms in hate speech and incitement to violence. The format appears to have enabled more frank discussion of internal policies and challenges than would likely have been possible in a public setting. In the context of AI, Council members could convene a similar meeting with representatives of frontier AI companies to discuss emerging risks, including to cybersecurity and information integrity, and to better understand the safety and risk-management measures they are putting in place. These formats could help move Council engagement beyond broad discussions of AI towards consideration of specific risks, their implications for particular situations, and possible opportunities for prevention or de-escalation.

Aide-Memoire on Technology

The Secretariat could also support continuity by developing an aide memoire on technology. Similar to the Protection of Civilians aide memoire, it could consolidate existing Council language on AI and other technologies by thematic issue and country situation. This would provide members with a common reference point for briefings, negotiations, and mandate renewals, while helping identify where agreed language already exists and where analytical gaps remain.

Conclusion

AI is already reshaping the geopolitical and security environment in which the Security Council operates. It is creating new opportunities for prevention, mediation, and peace operations, while amplifying existing risks in cyberspace, information environments, and armed conflict, and introducing more uncertain challenges as systems become increasingly capable and autonomous. The Council's engagement, however, remains uneven and reactive. More systematic mainstreaming and horizon-scanning, including through informal expert-level discussions, offer practical ways to strengthen its response. This approach could help the Council anticipate emerging risks, respond more effectively where AI affects situations within its remit, and identify ways AI can support international peace and security.

Security Council Report Staff

Shamala Kandiah Thompson
Executive Director

Katerina Limenopoulou
Chief Operating Officer

Paul Romita
Managing Editor

Alina Entelis
Deputy Managing Editor

Dawit Yirga Woldegerima
Deputy Managing Editor

Vladimir Sesar
Development and Outreach Manager

Jonathan Ahn
Operations Coordinator

Sara Bertotti
Senior Policy Analyst

Erik Ramberg
Senior Policy Analyst

Andrew Azzopardi
Policy Analyst

Joshua Levkowitz
Policy Analyst

Gaurav Redhal
Policy Analyst

Rodrigo Saad (lead author)
Policy Analyst

Security Council Report is a non-profit organisation supported by the Governments of Australia, Austria, Bahrain, Andrew Carnegie Foundation (formerly Carnegie Corporation of New York), Cyprus, Denmark, Finland, Friedrich-Ebert-Stiftung, Germany, Ireland, Konrad-Adenauer-Stiftung, Kuwait, Latvia, Liechtenstein, Luxembourg, Malta, the Netherlands, Norway, Open Society Foundations, Portugal, Republic of Korea, Singapore, Spain, Switzerland, Türkiye, United Arab Emirates, and Individual Contributions.

Design Point Five, NY

Security Council Report
125 Park Avenue
New York NY 10017

Telephone +1 212 759 9429
Web securitycouncilreport.org