14. Artificial intelligence and international peace and security

LAURA BRUUN, JULES PALAYER AND VINCENT BOULANIN

I. Introduction

The realization that artificial intelligence (AI)—and its implications for peace and security—is a multifaceted issue that requires a multifaceted response was evident in 2025. It was reflected in AI's appearance on agendas across intergovernmental processes and forums, both within and outside the United Nations. As such, global efforts to address the challenges and opportunities that AI presents for international peace and security increased in breadth and depth during the year. Discussions on autonomous weapon systems (AWS) expanded from Geneva to New York, the secretary-general of the United Nations published his first-ever report on AI in the military domain, and the UN General Assembly agreed to convene informal exchanges in 2026 on the peace and security implications of military AI. Efforts to understand the implications of civilian AI for peace and security also continued to evolve, for example, with UN member states agreeing to establish an Independent International Scientific Panel on AI (IISPAI). That said, many of the conversations on the peace and security implications of AI have remained at the exploratory stage—focusing on agreeing terminology and identifying key concerns, rather than, for example, devising concrete means to address those implications.

This chapter provides an overview of key efforts to understand and address challenges and opportunities posed by AI for international peace and security in 2025. It discusses, in turn, efforts that seek to address concerns around the military uses of AI (section II) and efforts that seek to address the risks that civilian AI presents to international peace and security (sections III). Section IV draws some brief conclusions. While there are many different definitions of AI and different technical understandings of what amounts to an AI system, AI is understood here in its broadest sense: as a catch-all term for computational techniques that allow computers and robots to solve complex, seemingly abstract problems that previously yielded only to human cognition. Examples include observing, navigating and making sense of the

world through vision; recognizing, processing and generating natural language; and learning how to control actuators to do multi-step tasks.[1]

II. Governing artificial intelligence in the military domain

International policy efforts to address the challenges presented by military AI and emerging technologies have historically been discussed through the lens of autonomous weapon systems. However, since 2023 policymakers have been dedicating increasing attention to the specific issue of AI in the military domain, notably in the context of the United Nations General Assembly and Security Council and as part of the Responsible AI in the Military Domain (REAIM) summit series (although the third summit was postponed from September 2025 to February 2026). This section outlines the main developments concerning the governance of AWS, as well as broader applications of military AI that took place in 2025.

Autonomous weapon systems

Meetings of the group of governmental experts under the Certain Conventional Weapons Convention

The challenges posed by AWS—commonly characterized as weapon systems that, once activated, can identify, select and engage targets without human intervention—have been the subject of intergovernmental discussions for more than a decade.[2] Technically framed around 'lethal' AWS (LAWS), these discussions have primarily taken place under the auspices of the 1981 Convention on Certain Conventional Weapons (CCW Convention), where a group of governmental experts (GGE) has led the work since 2017.[3] The overarching question guiding the work of the GGE is whether new regulation (e.g. in the form of a politically binding agreement or a new, legally binding protocol to the CCW Convention) is needed to address the challenges posed by LAWS.

More than 10 years into the discussions, the GGE on LAWS has established consensus around core principles. Among these are the principles that international humanitarian law applies to LAWS and that 'Human responsibility for decisions on the use of weapons systems must be retained since

[1] Boulanin, V. et al. *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, 2020).

[2] See the earlier discussions on the regulation of AWS in the 2014 and 2017–25 editions of the SIPRI Yearbook.

[3] For a summary and other details of the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects (CCW Convention) see annex A, section I, in this volume. On other developments in the CCW Convention see chapter 13, section II, in this volume.

accountability cannot be transferred to machines'.[4] There is also growing support for the governance of LAWS through a so-called two-tiered structure, which, in the first tier, would prohibit LAWS that cannot be used in compliance with international humanitarian law and, in the second tier, would place limits and requirements on all other LAWS.[5] However, which types and which uses of LAWS should be prohibited and which should be permitted but restricted remains subject to debate.

The GGE on LAWS has a three-year mandate to 'consider and formulate, by consensus, a set of elements of an instrument, without prejudging its nature' (notably whether or not it should be legally binding) and 'other possible measures to address emerging technologies in the area of lethal autonomous weapon systems' in 2024–26.[6] While the GGE was not required to submit a final report before the end of its mandate, 2025 provided an opportunity to discuss substantive issues in greater depth and identify areas of common ground. The group met for a total of 10 days (5 days in March and 5 in September) and, as in 2024, its meetings were structured around a reading of the 'rolling text'—a constantly updated document introduced by the chair of the GGE, Robert in den Bosch of the Netherlands, in 2024.[7] It is structured around five thematic sections: (I) working characterization of LAWS; (II) the application of international humanitarian law; (III) prohibitions and regulations; (IV) other measures to ensure compliance with international humanitarian law, and (V) accountability and responsibility.

The GGE completed several readings of the rolling text during its meetings in 2025, aiming to establish agreed language and shared understandings on key issues and concepts. Two themes received particular attention: characterizations of LAWS; and the term 'context-appropriate human control and judgment'.

Characterization. In the absence of a consensus definition of LAWS, states engaged in lengthy discussions in an effort to agree on a 'working characterization'. As in previous years, they expressed divergent views on several key points.

First, states debated whether to include the qualifier 'lethal' in the characterization of AWS. A few (e.g. China, Israel, Russia and the United States) insisted on retaining it, referring to the fact that 'lethal' is in the GGE's mandate and title. Most states argued that it should be deleted and that the GGE

[4] CCW Convention, GGE on LAWS, Report of the 2019 session, CCW/GGE.1/2019/3, 25 Sep. 2019, annex IV, paras a–b.

[5] CCW Convention, GGE on LAWS, Report of the 2023 session, CCW/GGE.1/2023/2, 24 May 2023; and Bruun, L., *Towards a Two-tiered Approach to Regulation of Autonomous Weapon Systems: Identifying Pathways and Possible Elements* (SIPRI: Stockholm, 2024).

[6] This and the full mandate of the GGE appear in CCW Convention, Meeting of the High Contracting Parties, Final report, CCW/MSP/2023/7, 23 Nov. 2023, para. 20.

[7] CCW Convention, GGE on LAWS, Rolling text, 18 Dec. 2025.

should address the challenges posed by all types of AWS, not only those with lethal effects. They argued, for example, that lethality is not an intrinsic characteristic of a weapon system but an effect of the use, and they further cautioned that its retention of the term 'lethal' may risk creating a loophole for AWS that cause injury to persons or damage to objects. To circumvent this disagreement, the GGE explored the possibility of additional text to clarify that LAWS without lethal effects or which can be used in circumstances that do not result in death are not excluded from the characterization. This proposal was based on a suggestion by the International Committee of the Red Cross (ICRC) and received widespread support, including initial support from such states as Israel and the USA.[8]

Second, states had extensive discussions about the nature and understanding of the key autonomous functions of an AWS: identifying possible targets, selecting a specific target and engaging that target. Some states (e.g. India, Japan and Singapore) favoured a cumulative approach, referring to AWS with all three capabilities. These states see identification as a precondition for selection and further note that identification is a key requirement for upholding the principle of distinction. They thus caution that the omission of identification risks unduly broadening the scope of the GGE to weapon systems outside its mandate (e.g. guided missiles). However, most other states (including New Zealand, Panama, Switzerland and the United Kingdom) argued that selection encompasses the task of identification and that inclusion of the term 'identify' may overly narrow the scope of their work.

These seemingly technical disagreements around characterization and definitions reflect a deeper disagreement that flows from states' vested interests in how broad or narrow the work of the GGE should be; that is, which types of weapon system with autonomous functions should be excluded from a future instrument.

'Context-appropriate human control and judgment'. The use of AWS raises novel questions about what compliance with international law requires from humans and permits from machines.[9] As a result, the question of how to capture in the work of the GGE on LAWS the human element in the use of force has been at the centre of the debate over the years. For example, some states have argued that compliance with international humanitarian law requires

[8] Varella, L., Reaching Critical Will, 'Editorial: Make it or break it', *CCW Report*, 11 Sep. 2025, p. 2.

[9] Bruun, L., Bo, M. and Goussac, N., *Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What does IHL Permit, Prohibit and Require?* (SIPRI: Stockholm, Mar. 2023).

'meaningful human control' over the use of AWS.[10] This view implies that the intervention or control by a human requires constant, analytical decision making. Meanwhile, others have maintained that the type of 'human involvement' required to comply with international humanitarian law ultimately depends on the context of use.[11]

After years of discussions, in 2025 there was growing support for the term 'context-appropriate human control and judgment'. This term, which was introduced by the chair into the rolling text, tries to bridge divergences on terminology and on the level of human control required by international law, particularly international humanitarian law. Most states endorsed the term, appreciating that it includes both 'control' and 'judgment' as these serve two different but equally important functions that are essential to ensure compliance with international humanitarian law.[12] Most also agreed that the level of control needed to use LAWS lawfully depends on the context of use, which is already recognized as a guiding principle.[13]

However, states were divided as to whether context-appropriate human control and judgment is an (implicit) requirement of international humanitarian law or a measure to comply with international humanitarian law.[14] States such as Austria and Pakistan, as well as the ICRC, argued that human control and judgment are implicit requirements of international humanitarian law. In contrast, states such as the USA argued that, while judgment and control offer useful guidance on how to ensure compliance with international humanitarian law, they are not existing legal requirements. Also, while the term generally received broad support, some states (e.g. India and Israel) had reservations. For example, they questioned whether the term would create uncertainties as to whether such control and judgment are required during all of the design, deployment and use phases.

Since the GGE's mandate concludes in 2026, states also addressed the way forward. During the September meeting, Brazil delivered a joint statement on behalf of 42 CCW states parties, representing a regionally and politically diverse range of states, expressing their readiness and desire to move towards

[10] E.g. CCW Convention, GGE on LAWS, 'Draft Protocol on Autonomous Weapon Systems (Protocol VI)', Working paper submitted by Argentina, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Kazakhstan, Nigeria, Palestine, Panama, Peru, the Philippines, Sierra Leone and Uruguay, CCW/GGE.1/2023/WP.6, 11 May 2023; and CCW Convention, GGE on LAWS, 'State of Palestine's proposal for the normative and operational framework on autonomous weapons systems', Working paper submitted by Palestine, CCW/GGE.1/2023/WP.2, 3 Mar. 2023.

[11] E.g. CCW Convention, GGE on LAWS, 'Working paper on the "human element"', Working paper submitted by Bulgaria, Denmark, Finland, France, Germany, Italy, Luxembourg, the Netherlands, Norway, Spain and Sweden, CCW/GGE.1/2025/WP.4, 21 Aug. 2025.

[12] CCW Convention, CCW/GGE.1/2025/WP.4 (note 11).

[13] CCW Convention, CCW/GGE.1/2019/3 (note 4), annex IV, para. c.

[14] Varella, L., Reaching Critical Will, 'Report of the meeting of the group of governmental experts on lethal autonomous weapons systems', *CCW Report*, 11 Mar. 2025, pp. 8–11.

negotiations on a new instrument within the CCW Convention (without specifying what form it would take and whether or not it should be legally binding), using the elements of the rolling text as the basis.[15] Other states argued that existing law is sufficient to address the challenges posed by AWS, and they do not see the need for the GGE to move towards adopting such a negotiating mandate. These states noted that the group should instead focus on continued technical exchanges and work towards further clarifying how existing international humanitarian law applies to AWS.

Overall, while 2025 was characterized by focused and constructive exchanges in the GGE on LAWS, deep divergences remain, including on key issues around the scope of the group's work, the limits and requirements that international humanitarian law places on the development and use of AWS and whether additional regulations are needed, and what would be the appropriate next steps for the GGE. It thus remains unclear whether agreement can be reached on the elements of a new instrument and whether the GGE will be able to reflect such agreement in its final report by the end of its mandate, in 2026. For now, most states continue to consider the CCW Convention to be the most appropriate forum to address the challenges posed by LAWS. However, it is important to note that states could also take the process to other another forum, either the UN General Assembly or—following the examples of the 1997 Anti-Personnel Mine (APM) Convention and the 2008 Convention on Cluster Munitions (CCM)—outside the UN entirely.[16] The attractiveness of these options will depend on the GGE's ability to reach consensus on the way forward by the end of its mandate in 2026.

Informal consultations on autonomous weapon systems at the United Nations in New York

On 12–13 May 2025 the UN General Assembly in New York hosted two days of informal consultations on LAWS, which was used as an opportunity to discuss AWS more broadly. The consultations, pursuant to UN General Assembly Resolution 79/62 from 2024, marked the first time that AWS were the subject of a mandated meeting held by the United Nations at its headquarters in New York, rather than in Geneva.[17] According to the resolution,

[15] CCW Convention, GGE on LAWS, Joint statement on behalf of Angola, Austria, Belgium, Brazil, Bulgaria, Chile, Colombia, Costa Rica, Denmark, the Dominican Republic, Ecuador, El Salvador, Finland, France, Germany, Guatemala, Hungary, Iceland, Ireland, Italy, Kazakhstan, Lesotho, Luxembourg, Mexico, Montenegro, Mozambique, Nauru, New Zealand, Nigeria, North Macedonia, Norway, Pakistan, Palestine, Panama, Peru, Portugal, Sierra Leone, Slovenia, Spain, Sweden, Switzerland and Uruguay, Sep. 2025. Two observer states also endorsed the statement: Kiribati and Thailand. The statement was repeated in CCW Convention, Meeting of the High Contracting Parties, Working paper submitted by Brazil et al., CCW/MSP/2025/WP.5, 26 Nov. 2025.

[16] For a summary and other details of the APM Convention and the CCM see annex A, section I, in this volume. On developments in these two conventions see chapter 13, sections III and V, in this volume.

[17] UN General Assembly Resolution 79/62, 'Lethal autonomous weapons systems', 2 Dec. 2024.

the purposes of the consultations included giving states an opportunity to consider the UN secretary-general's 2024 report on LAWS, to complement and support the GGE.[18]

A total of 96 states and several non-state actors from civil society, international organizations and academia participated in the consultations. The meeting opened with statements from, among others, the UN secretary-general and the president of the ICRC, who both called for the adoption of a legally binding instrument on AWS.[19] The agenda included a briefing by the chair of the GGE on LAWS, as well as a series of expert briefings on legal, humanitarian, security, technical and ethical considerations.[20] Following each briefing, states engaged in active discussions with the experts and with each other.

Since conversations on LAWS within the CCW framework are primarily focused on international humanitarian law, many UN member states and observers considered the consultations at the General Assembly a welcome opportunity to discuss the human rights implications of the use of LAWS outside armed conflicts (e.g. in law enforcement), as well as security challenges such as the proliferation of these weapons to non-state actors. The informal consultations triggered the recurring question of whether the CCW regime is the most suitable forum to address the challenges posed by AWS. To some, the appeal of the UN General Assembly is that it is more inclusive (since all states maintain a representation in New York, while not all are represented in Geneva, which hosts meetings under the CCW Convention) and its decisions do not require consensus. Nonetheless, most states expressed continued support for the CCW framework as the most suitable forum (for now) and held the view that conversations in New York should be seen as complementing, not replacing, efforts in Geneva, especially as the GGE was heading into the second half of its current mandate.

The UN General Assembly resolution on lethal autonomous weapon systems

In December 2025 the United Nations General Assembly adopted its third resolution on LAWS.[21] Like the two preceding resolutions, adopted in 2023 and 2024, this one was sponsored by Austria together with a diverse group

[18] UN General Assembly Resolution 79/62 (note 17), paras 5, 7.

[19] United Nations, 'Secretary-General's video message to the informal consultations on lethal autonomous weapons systems', 12 May 2025; and Spoljaric, M., ICRC president, 'Preserving human control over the use of force: A call to regulate lethal autonomous weapon systems under international law', International Committee of the Red Cross (ICRC), 12 May 2025.

[20] Open informal consultations on lethal autonomous weapons systems held in accordance with General Assembly Resolution 79/62, 'Programme', 30 Apr. 2025.

[21] UN General Assembly Resolution 80/57, 'Lethal autonomous weapons systems', 1 Dec. 2025.

of 30 co-sponsoring states.[22] The resolution was adopted with 164 states in favour, 6 against (Belarus, Burundi, Israel, North Korea, Russia and the USA), and 7 abstentions (Argentina, China, Iran, Nicaragua, Poland, Saudi Arabia and Türkiye).[23] While the USA had voted in favour of the 2024 resolution and Israel had abstained, both states voted against the 2025 text, a shift that is likely to be explained by, among other factors, the change of US administration and increased geopolitical tensions as well as specific national positions on the resolution's text.

The resolution expresses concern about the humanitarian, legal, security, technological and ethical challenges raised by AI and autonomy in weapon systems. It underlines the importance of the role of humans in the use of force to ensure responsibility and accountability and of compliance with international law. In its operative section, it calls on the states parties to the CCW Convention to 'work towards completing the set of elements for an instrument being developed within the mandate of the [GGE on LAWS], with a view to future negotiations'.[24] However, unlike the 2024 resolution on LAWS, which mandated two days of informal consultations in New York (described above), the 2025 resolution does not specify any further concrete activities but only encourages states to 'conduct further exchanges'.[25]

The fact that neither the informal consultations nor the resolution resulted in concrete steps suggests that states are awaiting the outcome of developments in 2026 in the GGE on LAWS before deciding on future action. Indeed, the current three-year mandate of the GGE has not only generated deep and focused discussions within the group but has resulted in states being cautious in other forums. Hence, there is growing pressure for the GGE to succeed. Undoubtedly, 2026 will be a defining year for the LAWS debate as future developments will hinge on whether and to what extent the GGE can reach consensus ahead of the 7th Review Conference of the CCW Convention, to be held in November 2026, as well as on the direction that the Review Conference ultimately sets for the GGE when determining its future mandate, if any.[26]

[22] UN General Assembly Resolution 78/241, 'Lethal autonomous weapons systems', 22 Dec. 2023; and UN General Assembly Resolution 79/62 (note 17). The main co-sponsoring states of the 2025 resolution were Austria, Belgium, Brazil, Costa Rica, Cuba, Guatemala, Ireland, Kazakhstan, Kiribati, Liechtenstein, Malta, Mexico, Moldova, New Zealand, the Philippines, San Marino, Sierra Leone, Sri Lanka, Switzerland, Tonga and Venezuela.

[23] United Nations, 'General Assembly adopts more than 60 resolutions, decisions of its First Committee (Disarmament and International Security)', GA/12736, 1 Dec. 2025.

[24] UN General Assembly Resolution 80/57 (note 21), para. 3.

[25] UN General Assembly Resolution 80/57 (note 21), para. 1.

[26] Blanchard, A. and Goussac, N., *Towards Multilateral Policy on Autonomous Weapon Systems* (SIPRI: Stockholm, Sep. 2025).

Military artificial intelligence

The UN secretary-general's report on artificial intelligence in the military domain

In June 2025 the UN secretary-general, António Guterres, published a report on AI in the military domain and its implications for international peace and security.[27] The report was the result of General Assembly Resolution 79/239 of 2024, which tasked the secretary-general to seek and summarize the views of UN member and observer states on the opportunities and challenges that the integration of AI in the military domain poses to international peace and security.[28] The resolution specifically tasked the secretary-general to seek views on the implications of AI in the military domain in areas other than LAWS. The secretary-general received a total of 49 submissions, of which 31 were from member states (with other submissions from the ICRC, civil society groups, the scientific community and industry). In comparison, when seeking views on LAWS in a similar exercise in 2024, the secretary-general received submissions from 73 states.[29] This suggests that states have a stronger capacity to address, or at least more developed positions on, the challenges posed by AWS.

The report on AI in the military domain summarizes key elements from the submissions received. These focus on (*a*) opportunities and challenges related to AI in the military domain; (*b*) a catalogue of existing and emerging normative proposals; (*c*) initiatives in the field of AI in the military domain; and (*d*) considerations on next steps. The report concludes with Guterres's observations and conclusions. These are summarized in turn below.

Among the key opportunities of AI in the military domain, states highlighted greater speed and scale, enhanced efficiency, improved accuracy and precision, and a lower probability of error compared to humans. They identified a wide range of opportunities for applications of AI, including intelligence, surveillance and reconnaissance (ISR) to support decision making and command and control; cyber defence and information and communications technology (ICT) security; peacekeeping; logistics; training and simulation; situational awareness; strategic analysis; and threat detection. According to some states, such applications highlight AI's potential to enhance operational effectiveness, civilian protection, and international peace and security. Meanwhile, key challenges highlighted in the report include shortening the time available for targeting decisions; the loss of human control, especially in

[27] United Nations, General Assembly, 'Artificial intelligence in the military domain and its implications for international peace and security', Report of the secretary-general, A/80/78, 5 June 2025.

[28] UN General Assembly Resolution 79/239, 'Artificial intelligence in the military domain and its implications for international peace and security', 24 Dec. 2024, paras 7–8.

[29] United Nations, General Assembly, 'Lethal autonomous weapons systems', Report of the secretary-general, A/79/88, 1 July 2024.

the context of the use of force; the potential for malicious use; the deepening technological asymmetries between states; and the risk of miscalculation and unintended escalation. In this regard, many contributions mentioned that these risks are particularly present when using AI to support use-of-force decisions and other combat-related functions.

Among the normative proposals, many states called for governance frameworks that are flexible, balanced, realistic and precautionary, focused on the entire life cycle of AI (including design, development and procurement), and based on the application and use of AI, rather than on the technology itself. From a technological governance perspective, states highlighted the need to incorporate guardrails to minimize harm; measures to prevent unintended consequences and malfunctions; clear operational boundaries and well-defined use cases; bias mitigation; and explainability, understandability, traceability and transparency.

Regarding initiatives in the field of AI in the military domain, the report summarizes all relevant UN-led processes.[30] It also highlights state-led initiatives, including the US-led Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy; and the REAIM process, initiated by the Netherlands and the Republic of Korea (South Korea).[31] However, it also notes the view that, while useful, such initiatives outside the UN risk fragmentation and could undermine more inclusive multilateral efforts in this area.

Considering next steps, most UN member states and observers called for increased dialogue on AI in the military domain in order to develop a regulatory or governance framework to mitigate the risks. States identified 11 priorities for future dialogue: (*a*) ensuring compliance with international law, in particular international humanitarian law; (*b*) protecting human dignity and human rights; (*c*) seeking common understandings on definitions and terminology; (*d*) considering transparency and confidence-building

[30] These include UN General Assembly Resolution 79/1, 'Pact for the Future', 22 Sep. 2024, and its annex I, 'Global Digital Compact'; UN General Assembly Resolution 79/239 (note 28); United Nations, Security Council, 'Harnessing safe, inclusive, trustworthy AI for the maintenance of international peace and security', Arria formula meeting, UN Web TV, 4 Apr. 2025; United Nations, General Assembly, Report of the Disarmament Commission for 2025, A/80/42, 30 Apr. 2025, under the agenda item 'Recommendations on common understanding related to emerging technologies in the context of international security', para. 21; UN General Assembly resolutions 78/241 (note 22) and 79/62 (note 17); and the work of the GGE on LAWS.

[31] Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, US Department of State, 9 Nov. 2023; and South Korean Ministry of Foreign Affairs, 'The Responsible AI in the Military Domain (REAIM) Summit 2024 opening session', 9 Sep. 2024. Other processes mentioned in the report include the Paris Declaration on Maintaining Human Control in AI enabled Weapon Systems, adopted at the AI Action Summit, 11 Feb. 2025; the Bletchley Declaration, adopted at the AI Safety Summit, 1–2 Nov. 2023; the work on AI taking place in the Group of 7 (G7), e.g. G7 Leaders' Statement on AI for Prosperity, 17 June 2025; and the Chinese-led Global AI Governance Initiative, Chinese Cyberspace Administration, 18 Oct. 2023, with an English translation by the Chinese Ministry of Foreign Affairs, 20 Oct. 2023.

measures; (*e*) addressing autonomy in the use of force; (*f*) addressing AI systems that directly support combat operations; (*g*) ensuring adequate data-governance mechanisms; (*h*) strengthening international cooperation and assistance; (*i*) supporting capacity-building and the sharing of good practices to bridge the digital divide; (*j*) promoting continued regional dialogue; and (*k*) promoting national regulation, including to ensure private sector compliance with international law.

The report closes with the secretary-general's observations and conclusions. These stress core concerns around loss of human control; the role of AI in facilitating hostilities in densely populated areas; the maintenance of human responsibility and accountability; and uncertainty on how international law applies to AI. In addition, Guterres urged all states that possess nuclear weapons to agree that any decision on nuclear use would be made by humans, not machines. He encouraged states to consider elaborating transparency and confidence-building measures at the regional and subregional levels, tailored to the unique characteristics and challenges of AI. Finally, the secretary-general recommended that states 'take concrete steps with a view to the establishment of a dedicated and inclusive process to comprehensively tackle the issue of AI in the military domain and its implications for international peace and security'.[32]

Overall, the report, as the first of its kind, provides a comprehensive baseline for states and other stakeholders to advance a still-emerging policy conversation on military AI. Reflecting this, just five months after the report's release, the UN General Assembly decided to convene informal exchanges on military AI in Geneva in 2026 (see below), a decision that aligns with the secretary-general's recommendations.

UN General Assembly resolution on artificial intelligence in the military domain

In December 2025 the UN General Assembly adopted its second resolution on 'Artificial intelligence in the military domain and its implications for international peace and security'.[33] The resolution, sponsored by the Netherlands and South Korea and co-sponsored by a further 66 states, was adopted with 167 in favour, 5 against (Burundi, Israel, North Korea, Russia and the USA) and 5 abstentions (Argentina, Belarus, Iran, Nicaragua and Saudi

[32] United Nations, A/80/78 (note 27), para. 66.

[33] UN General Assembly Resolution 80/58, 'Artificial intelligence in the military domain and its implications for international peace and security', 1 Dec. 2025.

Arabia).[34] The resolution builds on a resolution from 2024, also sponsored by the Netherlands and South Korea.[35] However, while the 2024 resolution tasked the secretary-general to seek the views of states on AI in the military domain, the operational section of the 2025 resolution contains a decision to 'convene informal exchanges in Geneva for States to further share perspectives on observations and conclusions as presented in the report of the Secretary-General'. These informal exchanges were scheduled to be held in June 2026, to last for three days and to be open to the full participation of all UN member and observer states, international and regional organizations, the ICRC, academia and civil society, including the scientific and technical community and industry.

The decision to convene informal exchanges can be seen as a step towards establishing a dedicated and inclusive process to tackle the issue of AI in the military domain, as requested by the secretary-general in his report. The exchanges can serve as an opportunity for states to elaborate on what types of AI application in the military domain deserve particular attention, to identify key challenges and to discuss possible avenues for the UN to formally address military AI. The fact that states decided to convene the exchanges in Geneva, rather than New York, suggests a wish to make use of the expert knowledge on military AI that the decade-long AWS debate has generated there. Geneva-based diplomats may also be better equipped to ensure complementarity, rather than duplication, between policy processes related to AWS and broader applications of military AI. However, holding the informal exchanges in Geneva does not exclude the possibility of future efforts taking place in New York, both places or elsewhere.

UN Security Council discussions on military artificial intelligence

In September 2025 South Korea convened a high-level open debate on AI and international peace and security in the UN Security Council. The Security Council had previously held two formal meetings on AI and implications for peace and security, one convened by the United Kingdom in 2023 and one by the United States in 2024. According to South Korea, a non-permanent member of the Council for the 2024–25 term, the latest meeting aimed to generate a discussion on how to maximize the benefits of AI while mitigating

[34] United Nations, GA/12736 (note 23). The main co-sponsoring states were Albania, Australia, Austria, Belgium, Bolivia, Bosnia and Herzegovina, Brazil, Bulgaria, Canada, Chile, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Iceland, Ireland, Italy, Kenya, South Korea, Latvia, Lithuania, Luxembourg, Malta, Moldova, Morocco, the Netherlands, Nigeria, North Macedonia, Norway, Pakistan, the Philippines, Poland, Portugal, Romania, San Marino, Singapore, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye, Ukraine and the UK.

[35] UN General Assembly Resolution 79/239 (note 28).

the risks, with a particular focus on the role of the Security Council.[36] Thus, without aiming to reach any decisions, the meeting served as an opportunity for states to exchange views on AI's implications on peace and security. Among other things, they considered the risks of misuse and loss of human control and the importance of using AI to promote peace and the common good.[37]

South Korea also used the meeting to launch the report *Responsible by Design* by the Global Commission on Responsible AI in the Military Domain (GC REAIM).[38] This commission is an independent body tasked by the 2023 REAIM Summit to produce 'a strategic guidance report which identifies short and long-term recommendations for governments and the wider multi-stakeholder community on responsible AI in the military domain'.[39] The launch of the GC REAIM report provided an opportunity for the REAIM process to maintain its activities, despite the postponement of the third REAIM Summit to February 2026.

III. Governing civilian artificial intelligence

In addition to efforts to govern the challenges of military AI, in 2025 states continued efforts to address and mitigate security-related challenges posed by civilian deployment of AI. Different initiatives both within and outside the United Nations sought to advance AI governance for the benefit of all while reducing the variety of risks that this technology poses. This section reviews these developments, starting with the initiatives that took place within the UN. Then, with a focus on efforts by states outside the UN, it looks at the AI Action Summit and the action taken by the European Union (EU) and by the two leading AI powers, China and the USA.

Efforts by the United Nations on global governance of artificial intelligence

In 2025 the UN General Assembly discussed the practical implementation of two key points of the Global Digital Compact, which had been adopted by the

[36] United Nations, Security Council, 'Concept note for the Security Council high-level open debate on the theme "Artificial intelligence and international peace and security: Addressing the complexities, multifaceted impacts and responsible use" in connection with the item "Maintenance of international peace and security", to be held on 24 September 2025', annex to S/2025/593, 22 Sep. 2025.

[37] United Nations, '"Innovation must serve humanity—not undermine it", cautions secretary-general, at Security Council open debate on artificial intelligence', SC/16180, 24 Sep. 2025.

[38] Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM), *Responsible by Design: Strategic Guidance Report on the Risks, Opportunities, and Governance of Artificial Intelligence in the Military Domain* (The Hague Centre for Strategic Studies: The Hague, Sep. 2025).

[39] The Hague Centre for Strategic Studies, 'Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM): Activities', [n.d.].

Summit of the Future in 2024: establishment of an Independent International Scientific Panel on AI and the creation of a multi-stakeholder Global Dialogue on AI Governance.[40] Costa Rica and Spain, which the General Assembly had appointed as facilitators of this process, led the discussion, which resulted in the adoption—without a vote—of the terms of reference and modalities for these new bodies.[41] In contrast to the developments discussed in section II, these two initiatives are mandated to focus on non-military uses of AI: within the UN, there is a clear separation between the governance of military uses of AI and the global governance challenges posed by civilian AI.

IISPAI is tasked with producing scientific and evidence-based reports on the 'opportunities, risks and impacts of artificial intelligence'.[42] The creation of this panel reflects the need to bring scientific perspectives into the discussions on the governance of AI, in particular as the conversation is too easily guided by the powerful states and corporations that have currently taken the lead in AI development.[43] Although the panel's reports will not be prescriptive, their conclusions are meant to be policy-relevant and to provide states with a rigorous, independent and multidisciplinary picture of AI opportunities, risks and impacts. IISPAI will have 40 members, with a broad geographical representation. Panel members were recommended by the UN secretary-general and appointed by the General Assembly early in 2026.[44] To further ensure equal representation, the panel will elect two co-chairs, 'one from a developed country and one from a developing country'.[45]

The Global Dialogue on AI Governance was launched in September 2025 in New York ahead of the new session of the General Assembly. Its multi-stakeholder framework is intended to ensure, as the UN secretary-general put it, that 'every country will have a seat at the table of AI'.[46] It will convene governments and other relevant stakeholders to 'discuss international cooperation, share best practices and lessons learned, and to facilitate open, transparent and inclusive discussions on artificial intelligence governance with a view to enabling artificial intelligence to contribute to the implementation of the Sustainable Development Goals [(SDGs)] and to closing the

[40] UN General Assembly Resolution 79/1 (note 30), annex I, para. 56. See also Palayer, J. and Bruun, L., 'Artificial intelligence and international peace and security', *SIPRI Yearbook 2025*, pp. 339–40.

[41] UN General Assembly Resolution 79/325, 'Terms of reference and modalities for the establishment and functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance', 26 Aug. 2025. See also United Nations, 'Secretary-general's remarks at high-level multi-stakeholder informal meeting to launch the Global Dialogue on Artificial Intelligence Governance', 25 Sep. 2025.

[42] UN General Assembly Resolution 79/325 (note 41), para. 1(a).

[43] Csernatoni, R., 'Corporate geopolitics: When billionaires rival states', Carnegie Endowment for International Peace, 30 Oct. 2025.

[44] United Nations, General Assembly, 'Appointment of members of the Independent International Scientific Panel on Artificial Intelligence', Note by the secretary-general, A/80/619, 3 Feb. 2026; and UN General Assembly Resolution 79/325 (note 41), para. 3.

[45] UN General Assembly Resolution 79/325 (note 41), para. 1(b).

[46] United Nations (note 41).

digital divides between and within countries'.[47] As an ambitious, global forum, the Global Dialogue will address a wide range of issues, including AI capacity gaps; open-source models; transparency; the development of safe, secure and trustworthy AI systems; the interoperability of governance approaches; and the assurance of human oversight over AI in compliance with international law.[48]

Overall, states expressed support for advancing multilateral engagement and cooperation on AI in the Global Dialogue on AI Governance. The USA was an exception: it did not attend the launch event and had rejected the initiative in the UN Security Council's high-level open debate the day before. It argued that any centralized control and global governance would stifle innovation and empower authoritarian uses of the technology (as discussed further below).[49] Future meetings of the Global Dialogue will be convened on the margins of relevant UN conferences and meetings, starting with the International Telecommunication Union (ITU) AI for Good summit in Geneva in July 2026.[50]

The AI Action Summit

The AI Action Summit was organized in Paris on 10–11 February 2025 by France in collaboration with India. The event served as a continuation of the AI Safety Summit series initiated by the UK at Bletchley Park in November 2023 and continued by South Korea in Seoul in May 2024.[51] The rebranding to 'AI Action' reflected the organizers' intent to broaden the agenda beyond a narrow focus on AI safety and risks associated with frontier AI—that is, the most advanced and general-purpose AI systems that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advance AI models—particularly the catastrophic risks associated with artificial general intelligence (AGI).[52] Instead, the 2025 summit focused on tangible applications that prioritize the benefits of AI and its potential to advance the SDGs. The summit was framed by three guiding questions: (*a*) how AI technologies and uses can be developed globally and deployed widely and equitably; (*b*) how inclusivity and freedoms can be preserved in

[47] UN General Assembly Resolution 79/325 (note 41), para. 4.

[48] Wilkinson, I., 'Can the UN's new AI governance efforts weather the AI race?', Chatham House, 18 Sep. 2025.

[49] US Mission to the UN, 'Remarks at the Security Council's open debate on artificial intelligence and international peace and security', 24 Sep. 2025; and Caroli, L. and Mande, M., 'What the UN Global Dialogue on AI Governance reveals about global power shifts', Critical Questions, Center for Strategic and International Studies (CSIS), 7 Oct. 2025.

[50] UN General Assembly Resolution 79/325 (note 41), paras 5–6.

[51] Bletchley Declaration (note 31); Boulanin, V., 'Governing the challenges presented by autonomous weapon systems', *SIPRI Yearbook 2024*, pp. 537–38; and Palayer and Bruun (note 40), pp. 340–42.

[52] British Department for Science, Innovation and Technology, 'Capabilities and risks from frontier AI: A discussion paper on the need for further research into AI risk', AI Safety Summit, Oct. 2023, p. 30.

the AI revolution; and (*c*) how AI can be used in ways that respect humanist values and serve society and the public interest.[53] While safety remained on the agenda, the topic was juxtaposed with consideration for trustworthiness, diversity and sustainability.

The reaction to this thematic shift was mixed. Stakeholders from academia, civil society and the private sector who had sought productive discussions and policy commitments on risks associated with advanced AI expressed disappointment.[54] For them, this change entailed a deprioritization of AI safety. As evidence of this, they noted that Yoshua Bengio, who had been commissioned to lead the production of a seminal report on AI safety—modelled on reports by the Intergovernmental Panel on Climate Change (IPCC)—was not invited to present the report's conclusions as part of the official programme (although he did discuss the report in side events).[55] Some also regretted that the programme did not provide a clear opportunity for AI companies to follow up on the commitments they had made at the 2024 Seoul Summit on development of the frontier safety frameworks—a set of structured internal protocols and policies for proactively identifying, evaluating and mitigating severe, large-scale risks posed by highly capable frontier AI models.[56] Others expressed concern that the final declaration did not include an ambitious commitment regarding AI safety.[57] Consequently, some civil society groups, such as the Future of Life Institute, called on states not to sign the declaration.[58]

In contrast, stakeholders historically critical of the focus on advanced AI risks welcomed the summit's change of direction.[59] They viewed the emphasis on diversity, inclusivity and sustainability as a positive development as it allowed discussion of near-term concerns, such as the energy consumption of large models and bias and the performance problems encountered outside the anglosphere. They also welcomed the announcement by the EU of

[53] AI Action Summit, 'Presentation', 17 Jan. 2025.

[54] Hashim, S., 'The embarrassing failure of the Paris AI summit', Transformer, 11 Feb. 2025; Kurbalija, J., 'The Paris AI Summit: A diplomatic failure or a strategic success?', Diplo, 12 Feb. 2025; and Perrigo, B., 'Safety takes a backseat at Paris AI Summit, as US pushes for less regulation', *Time*, 11 Feb. 2025.

[55] Bengio, Y. (chair), *International AI Safety Report: International Scientific Report on the Safety of Advanced AI*, British Department for Science, Innovation and Technology (DSIT) Research Series no 2025/001 (DSIT: London, Jan. 2025).

[56] British Department for Science, Innovation and Technology, 'Frontier AI safety commitments, AI Seoul Summit 2024', 7 Feb. 2025; and Frontier Model Forum, 'Components of frontier AI safety frameworks', Issue brief, 8 Nov. 2024.

[57] AI Action Summit, 'Statement on inclusive and sustainable artificial intelligence for people and the planet', 11 Feb. 2025.

[58] Tegmark, M. (@tegmark), X, 10 Feb. 2025, <https://x.com/tegmark/status/1888902965907370129>.

[59] Bryson, J., LinkedIn, 11 Feb. 2025, <https://www.linkedin.com/posts/bryson_aiactionsummit-sommetactionia-activity-7295370484479590400-W2fh>.

new public investment in AI research and infrastructure.[60] The necessity of ensuring that research is not solely funded by the private sector was further emphasized during the summit's high-level segment by Amandeep Singh Gill, the UN secretary-general's envoy on technology.

The final declaration of the AI Action Summit, which was signed by 62 states, the African Union and the EU, highlights the importance of reinforcing the diversity of the AI ecosystem and outlines six priorities for further work: (*a*) promoting AI accessibility to reduce digital divides; (*b*) ensuring that AI is open, inclusive, transparent, ethical, safe, secure and trustworthy, 'taking into account international frameworks for all'; (*c*) making innovation in AI thrive by enabling conditions for its development, avoiding market concentration, and driving industrial recovery and development; (*d*) encouraging AI deployment that positively shapes the future of work and labour markets and delivers opportunity for sustainable growth; (*e*) making AI sustainable for people and the planet; and (*f*) reinforcing international cooperation to promote coordination in international governance.[61]

While China signed the declaration, neither the United Kingdom nor the United States did (and Russia was not invited to the summit). The USA's decision not to sign was interpreted as a direct consequence of its administration's approach to AI, which emphasizes a deregulatory stance (see below).[62] The UK explained that it did not sign because of concerns about national security and 'global governance'.[63] Its decision surprised many stakeholders, not least because the UK initiated the summit series. Some commentators interpreted this decision as disapproval of the summit's thematic shift and as a diplomatic manoeuvre caused by the USA's new approach to AI governance.[64] In contrast, the facts that China signed the document and that India was to host the next summit in the series (the AI Impact Summit in February 2026) have been viewed as evidence that these countries intend to play a greater role in talks on AI governance.[65]

[60] European Commission, 'EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence', Press release, 11 Feb. 2025.

[61] AI Action Summit (note 57).

[62] Sanger, D. E., 'Vance, in first foreign speech, tells Europe that US will dominate AI', *New York Times*, 11 Feb. 2025.

[63] Kleinman, Z. and McMahon, L., 'UK and US refuse to sign international AI declaration', BBC, 11 Feb. 2025.

[64] Davies, P., ' "Devoid of any meaning": Why experts are calling the Paris AI Action Summit a "missed opportunity" ', Euronews, 14 Feb. 2025; and Klaus, J., 'Liberal democracies are retreating from AI safety', Lawfare, 11 Aug. 2025.

[65] 'China is leading the world on AI governance: Other countries must engage', *Nature*, 11 Dec. 2025; and Polcumpally, A. T., 'India's path to AI leadership: An alternative AI governance to US–China competition', Asia Society Policy Institute, [Oct. 2025].

The European Union's implementation of the AI Act and other strategies

Having adopted the AI Act in 2024, the EU continued its implementation in 2025, notably by releasing the Code of Practice for General-purpose AI Models for AI developers while also issuing key strategies for AI.[66]

The Code of Practice for General-purpose AI Models

The Code of Practice for General-purpose AI Models is a voluntary set of practices to help providers of general-purpose AI models to comply with the obligations of the AI Act.[67] It was the result of a year-long multi-stakeholder consultation that brought together a broad range of participants, and was drafted by independent experts appointed by the EU's European AI Office.[68] The Code of Practice has three chapters: the first two—on transparency and copyright—are for all providers of general-purpose AI models; the third—on safety and security—is only applicable to providers of general-purpose AI models with systemic risk.[69] According to the AI Act, 'systemic risk' means potential for 'large-scale harm from the most advanced . . . models' in the form of, for example, a loss of control over autonomous general-purpose AI models or the 'lowering of barriers for chemical or biological weapons development'.[70]

The main contribution of the Code of Practice is that it specifies what companies should consider when developing safe models. Developers are directed to assess four risks by default: (*a*) chemical, biological, radiological and nuclear (CBRN); (*b*) cyber offence; (*c*) loss-of-control risk—which means the risk of humans losing the ability to control, modify or shut down a model; and (*d*) harmful manipulation.[71] Major AI companies had already started considering these risks in the context of their internal safety-evaluation programmes, but not in a systematic or harmonized way. This makes the Code of Practice an important tool that provides the industry with clear guidance and a common framework to assess risks consistently when building AI

[66] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), *Official Journal of the European Union* L, 12 July 2024. On adoption of the act see European Parliament, 'Artificial Intelligence Act: MEPs adopt landmark law', Press release, 13 Mar. 2024; and Palayer and Bruun (note 40), pp. 342–43.

[67] European Commission, Directorate-General for Communications Networks, Content and Technology, 'The General-purpose AI Code of Practice', 10 Dec. 2025. See also European Commission, Directorate-General for Communications Networks, Content and Technology, 'Questions and answers on the Code of Practice for General-purpose AI', 11 July 2025.

[68] European Commission, 'Questions and answers' (note 67). See also Regulation (EU) 2024/1689 (note 66), para. 110ff.

[69] European Commission, 'Questions and answers' (note 67).

[70] European Commission, Directorate-General for Communications Networks, Content and Technology, 'General-purpose AI models in the AI Act—Questions & answers', 9 Sep. 2025.

[71] Papadatos, H., 'How the EU's Code of Practice advances AI safety', AI Frontiers, 12 July 2025.

models.[72] As of January 2026 the Code of Practice had been signed by most major AI companies (e.g. Amazon, Anthropic, Google, IBM, Microsoft, Mistral AI and OpenAI).[73] Notable exceptions were xAI, which only endorsed the chapter on safety and security; Meta, which refused to endorse the code amid the 'legal uncertainties' that it introduces; and Chinese companies, none of which had endorsed the code.[74]

The European Union's AI strategies

In 2025 the EU's European Commission released two key strategies related to AI: the AI Continent Strategy and the Apply AI Strategy.

The AI Continent Strategy, released in April 2025, sets out a path for the EU to achieve its goal of becoming a global leader in AI.[75] The strategy is built on the assumption that the 'race for leadership in AI is far from over' and sets five objectives necessary 'for Europe to become the AI Continent'. The first objective is to build and strengthen the EU's computing resources ecosystem by building collaborative AI hubs—'AI factories'—and investing in massive facilities—'AI gigafactories'—capable of training advanced models. The second is to ensure access to reliable, well-organized data, which is essential for unlocking the full potential of AI and supporting its development and innovation. Third is to foster innovation and boost deployment and use of AI across various sectors of the EU's economy. The fourth objective is to strengthen AI skills and talent in the EU by enlarging its pool of AI specialists and 'upskilling and reskilling the EU workforce'. Fifth is the fostering of regulatory compliance by streamlining and simplifying the application of the AI Act.

The European Commission released the Apply AI Strategy in October 2025.[76] Following the third objective of the AI Continent Strategy, it proposes that EU companies and the public sector adopt an 'AI first policy' approach to problem-solving. In other words, the Apply AI Strategy aims to boost the use of AI and facilitate its integration into day-to-day business operations. The strategy establishes three pillars: (*a*) introducing sectoral flagships 'to boost AI use in key industrial sectors of the EU economy'; (*b*) addressing cross-cutting challenges 'to scale the development and integration of AI across the EU strategic sectors'; and (*c*) establishing a single governance mechanism 'To

[72] Papadatos (note 71); and Boulanin, V., Palayer, J. and Ovink, C., *Addressing the Risks that Civilian AI Poses to International Peace and Security: The Role of Responsible Innovation* (SIPRI: Stockholm, 2025).

[73] European Commission, 'The General-purpose AI Code of Practice' (note 67).

[74] Henning, M., 'Meta won't sign EU's Code of Practice for generative AI', Euractiv, 18 July 2025.

[75] European Commission, 'AI Continent Action Plan', Communication to the European Parliament et al., COM(2025) 165 final, 9 Apr. 2025.

[76] European Commission, 'Apply AI Strategy', Communication to the European Parliament and the Council, COM(2025) 723 final, 8 Oct. 2025.

structure a continuous dialogue on AI' among key stakeholders and avoid a top-down governance model.

One notable aspect of the Apply AI strategy is that it includes the defence sector as one of 11 flagship sectors for the application of AI, dedicated to accelerating the development and deployment of EU AI-enabled situational-awareness and command-and-control capacities and to deploy a strategic EU infrastructure of highly secured computing power capacities. This is a departure from the traditional EU approach of separating the governance of military AI from that of civilian AI.

The year ended with the European Commission taking some steps to simplify the AI Act and slow down its implementation, in line with the fifth objective of the AI Continent Strategy, through the adoption of a 'digital omnibus'.[77] According to the Commission, this document addresses 'delays in designating national competent authorities and conformity assessment bodies, as well as a lack of harmonised standards for the AI Act's high-risk requirements, guidance, and compliance tools'.[78] Some commentators viewed these changes as being the result of 'intense lobbying' by the USA and from 'corporate lobbies' against the AI Act.[79]

Developments in China

In July 2025 China issued its Global AI Governance Action Plan, which builds on and expands its 2023 Global AI Governance Initiative.[80] The action plan is China's attempt to fulfil the commitments made in the Pact for the Future and the Global Digital Compact.[81] It outlines China's vision for global governance of AI, which is heavily reliant on multilateral cooperation, capacity-building and sharing. Indeed, most of the action plan focuses on the role of states and international organizations in fostering international cooperation, especially engaging with states outside the Global North. In this vein, in 2025 China proposed establishing a World Artificial Intelligence Cooperation Organization (WAICO), which is intended to reimagine the architecture of AI

[77] Haeck, P. and O'Regan, E., 'Brussels is done being the world's digital policeman', Politico, 25 Nov. 2025.

[78] European Commission, 'Digital omnibus on AI', Proposal for a regulation of the European Parliament and of the Council amending regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence, COM(2025) 836 final, 19 Nov. 2025.

[79] Haeck, P., 'The EU promised to lead on regulating artificial intelligence. Now it's hitting pause', Politico, 19 Nov. 2025.

[80] Chinese Ministry of Foreign Affairs, 'Global AI Governance Action Plan', 26 July 2025; and Global AI Governance Initiative (note 31).

[81] UN General Assembly Resolution 79/1 (note 30); and Chinese Ministry of Foreign Affairs (note 80).

governance, complementing big UN-led multinational initiatives and other efforts such as the AI Safety Summit series.[82]

Meanwhile, China focused on helping other countries to build their AI capabilities.[83] For example, China and Zambia hosted meetings of the Group of Friends for International Cooperation on AI Capacity-building, an initiative created by the two countries in December 2024 to foster 'policy exchanges, knowledge sharing, and practical collaboration'.[84] In 2025 China also took part in several initiatives that focus on AI-related capacity-building in the Global South hosted by BRICS, the League of Arab States, and the Community of Latin American and Caribbean States (Comunidad de Estados Latinoamericanos y Caribeños, CELAC).[85]

Nationally, China is making more specific calls to reduce AI risks, while still stressing that safety and development should go hand in hand.[86] It is putting national AI rules into practice through a growing system of technical standards 'on privacy, data security, AI-generated content labeling, harmful and illegal content filtering, and general security management practices'.[87]

Developments in the United States

The change of administration in 2025 in the United States had a notable impacted on AI policy. As soon as he took office, US President Donald J. Trump repealed the previous administration's executive order on safe, secure and trustworthy development and use of AI and issued a replacement a few days later.[88]

In July the new administration released an AI Action Plan that aims for the USA to 'achieve and maintain unquestioned and unchallenged global technological dominance' in AI.[89] The action plan focuses on three pillars: (*a*) accelerating AI innovation; (*b*) building US AI infrastructure; and

[82] Gibney, E., 'China wants to lead the world on AI regulation—Will the plan work?', *Nature*, 1 Dec. 2025; and Xu, Y., 'China's WAICO proposal and the reordering of global AI governance', CGTN, 30 July 2025.

[83] Wagner, G. et al., *State of AI Safety in China* (Concordia AI: July 2025).

[84] Chinese Permanent Mission to the UN, 'A side event of the Group of Friends for International Cooperation on AI Capacity-building successfully held at the UN headquarters', 6 May 2025; and Chinese Permanent Mission to the UN, 'Group of Friends for International Cooperation on AI Capacity-building formally established', 5 Dec. 2024.

[85] Wagner et al. (note 83), p. 27. For brief descriptions and lists of members of BRICS and the Arab League see annex B, sections I and II, in this volume.

[86] Wagner et al. (note 83), p. 7.

[87] Wagner et al. (note 83), p. 15.

[88] The White House, 'Removing barriers to American leadership in artificial intelligence', Executive Order no. 14 179, 23 Jan. 2025, *Federal Register*, 31 Jan. 2025; and The White House, 'Safe, secure, and trustworthy development and use of artificial intelligence', Executive Order no. 14 110, 30 Oct. 2023, *Federal Register*, 1 Nov. 2023.

[89] The White House, *Winning the AI Race: America's AI Action Plan* (The White House, Washington, DC, July 2025), p. i.

(*c*) leading in international AI diplomacy and security. Across the pillars, three core principles apply: AI developments need to benefit US workers and their families; 'AI systems must be free from ideological bias and be designed to pursue objective truth rather than social engineering agendas when users seek factual information or analysis'; and the misuse and theft of AI by malicious actors needs to be prevented and these risks should be monitored.[90]

The first policy action under the first pillar is to 'remove red tape and onerous regulation', which is justified on the basis that the previous administration's executive order 'foreshadowed an onerous regulatory regime'.[91] It then tasks different government offices to search for and revise or repeal this onerous regulation. Under this first pillar, the AI Action Plan also fosters adoption of AI in the public sector, including the Department of Defense (DOD), stating that the USA must 'aggressively adopt AI within its Armed Forces if it is to maintain its global military preeminence'.[92]

The policy action encapsulated in the second pillar covers the building of data centres and semiconductor manufacturing, as well as developing an energy grid to match the pace of AI innovation. Under this pillar, the US administration has vowed to reject 'radical climate dogma and bureaucratic red tape'.[93]

The third pillar is the most outward-looking. It aims to export US technology standards to allies while restricting competitors such as China, including by better enforcement of existing controls on advanced chips and by plugging loopholes in restrictions on exports of semiconductor-manufacturing equipment. Another key point under this pillar is to counter Chinese influence in international governance bodies. Hence, it also calls for the creation of an 'enduring global alliance' of countries using US-developed AI technology stacks to prevent international dependence on the technologies of the USA's adversaries.[94]

IV. Conclusions

The realization that the challenges—and opportunities—that AI presents for international peace and security are no longer a hypothetical issue was reflected clearly in the international policy landscape in 2025. Policy efforts to address military AI both broadened—with the topic of AI appearing on agendas across the United Nations—and deepened—not least due to the UN secretary-general's report on military AI.

[90] The White House (note 89), pp. 1–2.
[91] The White House (note 89), p. 3.
[92] The White House (note 89), p. 11.
[93] The White House (note 89), p. 1.
[94] The White House (note 89), p. 20.

However, apart from the topic of autonomous weapon systems (which has now been discussed for more than a decade), many of the conversations around AI in the military domain remained at a somewhat nascent or exploratory stage in 2025. As pointed out by the secretary-general, further dialogue is needed on several areas to advance understanding of the challenges and opportunities that AI presents for international peace and security. However, to advance such a dialogue, there is a need to strengthen capacity-building, political will and good-faith multilateralism. Considering the geopolitical climate in 2025, the realization of those elements in the near future is likely to prove difficult. That said, the policy conversations remain active and, with many states paying increasing attention to the implications of AI for peace and security, some developments in this field are to be expected.