12. Artificial intelligence and international peace and security

JULES PALAYER AND LAURA BRUNN

I. Introduction

For the past decade the international policy conversation on military uses of artificial intelligence (AI) has mostly focused on autonomous weapons systems (AWS), commonly characterized as weapon systems that, once activated, can select and engage targets without human intervention.[1] However, since 2023 the conversation has expanded to other military uses of AI, such as in targeting, planning and intelligence analysis, which are commonly referred to as AI-enabled decision support systems.[2] Reported uses of AI in current armed conflicts, especially in Gaza and Ukraine, have illustrated that the issue of military AI is a pressing matter for policy makers.[3]

Civilian AI developments can also pose risks to peace and security.[4] For example, some AI models could help malicious actors to access critical knowledge to develop and use prohibited weapons (i.e. chemical, biological, radiological and nuclear weapons).[5] In the same vein, AI provides a capability uplift and lowers the barrier for cybercriminal and hackers to carry out harmful operations.[6] Generative AI tools can be misused to spread

[1] Boulanin, V., 'Governing the impact of artificial intelligence on international peace and security', *SIPRI Yearbook 2024*, Oxford University Press, pp. 525–39.

[2] See e.g. Holland, A., *Decision, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-making* (International Committee of the Red Cross (ICRC): Geneva, 2024); and Nadibaidze, A., Bode, I. and Zhang, Q., *AI in Military Decision Support Systems: A Review of Developments and Debates*, AutoNorms project report (Center for War Studies, University of Southern Denmark: Odense, 2024).

[3] Franke, U. and Söderström, J., 'Star tech enterprise: Emerging technologies in Russia's war on Ukraine', European Council on Foreign Relations Policy Brief, 5 Sep. 2023; and McKernan, B. and Davies, H., ' "The machine did it coldly": Israel used AI to identify 37,000 Hamas targets', *The Guardian*, 3 Apr. 2024.

[4] Boulanin, V. and Ovink, C., 'Civilian AI is already being misused by the bad guys', *IEEE Spectrum*, 27 Aug. 2022.

[5] OpenAI, 'Preparedness framework (beta)', 18 Dec. 2023; Anthropic, 'Responsible scaling policy', 15 Oct. 2024; and Google Deepmind, 'Frontier safety framework: version 1.0', [n.d.].

[6] British National Cyber Security Centre, 'The near-term impact of AI on the cyber threat', 24 Jan. 2024.

disinformation, contributing to the erosion of the 'trust and belief in the broader informational and political environment'.[7]

This chapter takes stock of how states have tried to address these different concerns at the multilateral, regional, bilateral and national levels. It discusses, in turn, efforts that seek to address concerns around the military use of AI (section II) and those that seek to address the risks that civilian AI presents to international peace and security (sections III and IV). Section V draws some brief conclusions.

II. Governing the challenges presented by military artificial intelligence

The broadening of the policy conversation around military use of AI is reflected in the creation (or continuation) of new forums and initiatives. This section outlines the main developments concerning the governance of AWS as well as broader applications of military AI that took place in 2024.

Autonomous weapon systems

The 2024 meetings of the group of governmental experts at the CCW

The intergovernmental debate on AWS has since 2013 been discussed mainly through the lens of 'lethal autonomous weapon systems' (LAWS), notably within the 1981 Certain Conventional Weapons (CCW) Convention.[8] The discussions, which since 2017 have been led by a group of governmental experts (GGE), centre around whether the challenges posed by LAWS warrant the adoption of a new regulation, for instance, in the form of a new protocol under the CCW Convention. While this remains debated, growing support to govern LAWS through a so-called two-tiered structure has emerged over the past couple of years. The proposed two-tiered structure would prohibit certain types of LAWS and place limits and requirements on all other types of LAWS. However, what should be prohibited, restricted or required remains undecided. The GGE meetings in 2024 provided an opportunity for states to discuss the issue further.

The GGE on LAWS, chaired by Dutch ambassador Robert in den Bosch, has a three-year mandate to 'further consider and formulate, by consensus, a set of elements of an instrument, without prejudging its nature, and other

[7] Schiff, D. S., Jackson Schiff, K. and Bueno, N., 'Watch out for false claims of deepfakes, and actual deepfakes, this election year', Brookings Commentary, 30 May 2024; and Nimmo, B. and Flossman, M., *Influence and Cyber Operations: An Update* (OpenAI: San Francisco, Oct. 2024).

[8] 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 11, section II, in this volume.

possible measures to address emerging technologies in the area of LAWS'.[9] Given its three-year mandate, the GGE did not have to adopt a final report by the end of the year. Instead, 2024 provided an opportunity to discuss substantive issues in greater depth and identify areas of common ground for a substantial report in 2026. Structured around a 'rolling text' prepared by the chair, three topics were at the centre of 2024's discussions: characteristics and definitions of LAWS, application of international humanitarian law (IHL), and measures to ensure compliance with IHL and mitigate risks. The GGE discussed these topics during their two formal meetings in 2024 (one five-day session in March and one five-day session in August) and during a series of informal online intersessional consultations. The following presents some of the main contours of these discussions.

Definitions and characterizations. After more than a decade of discussions within the CCW, states have still not agreed on a definition of LAWS. For some states, such as the Russian Federation and Türkiye, agreeing on a definition is a prerequisite for any regulatory discussions.[10] Most other states are of the view that agreeing on a 'working characterization' suffices to identify elements of regulation. Following the latter approach, the GGE chair initiated lengthy discussions around such a working characterization in 2024. Most states supported the characterization of LAWS as 'weapon systems that, once activated, can select and engage a target without further intervention by a human operator'.[11] This characterization has already been used for years by many states and the International Committee of the Red Cross (ICRC).[12] However, the GGE was not able to reach an agreement about a working characterization due to enduring disagreements over how to reflect the human role in targeting decisions, what 'weapon systems' should be included and whether a characterization needs to include the 'lethal' qualifier. On the latter point, a majority held that 'AWS' is preferable to 'LAWS', many arguing that 'lethality' pertains to how the weapon system is used and its effects rather than the way it is designed, and that AWS are capable of causing harm in the form of material damage or injury, irrespective of whether death was the intended or actual result. That the GGE did not manage to establish a shared conceptual understanding of LAWS after more

[9] Meeting of the CCW High Contracting Parties, Final report, CCW/MSP/2023/7, 23 Nov. 2023, para. 20.

[10] Varella, L., 'Characterisation', *CCW Report*, vol. 11, no. 2 (14 Mar. 2023), p. 8.

[11] Varella, 'Characterisation' (note 10).

[12] See e.g. ICRC, 'ICRC Position on Autonomous Weapon Systems', 12 May 2022; and CCW Convention, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (GGE on LAWS), 'Draft articles on autonomous weapon systems—prohibitions and other regulatory measures on the basis of international humanitarian law ("IHL")', Working paper submitted by Australia, Canada, Japan, the Republic of Korea, the United Kingdom and the United States, CCW/GGE/1/2023/WP.4, 6 Mar. 2023, p. 1.

than a decade of discussions was, unsurprisingly, a great disappointment to many, both inside and outside of the group.

Compliance with IHL. The GGE previously established that IHL applies fully to the development and use of LAWS.[13] In 2024 the group discussed what specific prohibitions, limits and requirements flow from IHL, and how these could be reflected in a two-tiered instrument. For the GGE, it was beyond dispute that LAWS that cannot be used in compliance with IHL are prohibited. Also, except for a few states, notably Russia, most agreed that LAWS whose effects and behaviour cannot 'be reasonably anticipated nor controlled' should be considered prohibited because such weapon systems cannot be used in compliance with IHL. However, the GGE was divided as to whether LAWS whose effects and behaviour *also* cannot be explained and traced back (to responsible agents) would be off-limits as well. While the ICRC and several states, such as Austria, Pakistan, Mexico and Palestine, argued that such systems would contravene IHL compliance, others, such as the United States, disagreed, warning against conflating *compliance* with IHL with the *behaviours* that are required for compliance. In response, Pakistan argued that requirements around explainability and traceability have a basis in IHL but that they are things states have not been forced to think about before.[14] The GGE also discussed (without agreeing) specific limits that IHL places on the design and use of LAWS. A large group of states, including New Zealand, Pakistan and Palestine, as well as the ICRC, argued for the need to limit the use of LAWS to areas where civilians are not present and for LAWS only to be designed and used to target objects, not humans.[15] Other states, such as Germany, Israel and the United Kingdom, opposed such categorical limits, stressing that compliance with IHL depends on the context.[16]

Measures to ensure compliance with IHL. As in previous years, the GGE discussed the importance of rigorous testing and evaluation, legal reviews and training of users, with the aim of enhancing understanding of system capabilities and limitations in expected circumstances of use. In addition, the importance of ensuring measures to detect and reduce bias in the systems received more attention in 2024 than in previous years.[17] While there was

[13] CCW Convention, GGE on LAWS, Report of the 2019 session, CCW/GGE.1/2019/3, 25 Sep. 2019, annex IV, 'Guiding principles'.

[14] Acheson, R. and Varella, L., 'Topic 2: Application of international humanitarian law', *CCW Report*, vol. 12, no. 1 (12 Mar. 2024).

[15] CCW Convention, GGE on LAWS, 'Elements of an international legal instrument on lethal autonomous weapons systems (LAWS)', Working paper submitted by Pakistan, CCW/GGE.1/2024/WP.7, 14 May 2024, section IIC; Acheson and Varella (note 14), pp. 15, 17; and Varella, L. and Bjertén, E., 'Prohibitions and restrictions', *CCW Report*, vol. 12, no. 3 (5 Sep. 2024), p. 40.

[16] Acheson and Varella (note 14), pp. 15, 19 and 20.

[17] Varella and Bjertén (note 15), p 46; and CCW Convention, GGE on LAWS, 'Addressing bias in autonomous weapons', Working paper submitted by Canada, Costa Rica, Germany, Ireland, Mexico and Panama, CCW/GGE.1/2024/WP.5, 7 Mar. 2024.

broad support for the need to have various risk mitigation measures in place, states disagreed on whether such measures should be binding or voluntary, and thus how they should be reflected in a potential two-tiered regulation.

Overall, the 2024 CCW debate on LAWS was characterized by rich discussions and a constructive spirit among states about the need to identify compromise language that could lead to tangible outcomes. However, the positive tenor can largely be explained by the fact that the GGE did not have to produce and agree on a report by the end of the year. A final report is not due for adoption before 2026, and whether the substantive discussions will be reflected in such a consensus document is far from guaranteed. Underneath the constructive atmosphere remain fundamental disagreements, notably about whether a new legally binding instrument is needed and whether regulation should solely be grounded in IHL, as well as other concerns, especially those related to ethics and human rights. Nonetheless, 2024 concluded with a chair's summary and a revised version of the entire rolling text as a basis to support discussions in 2025.[18]

Regional and international conferences on AWS

Following last year's wave of regional conferences on AWS (in Costa Rica, Trinidad and Tobago, and the Philippines), Sierra Leone hosted a regional conference on AWS for members of the Economic Community of West African States (ECOWAS) on 17–18 April. The conference concluded with the adoption of a communiqué in which ECOWAS members stressed that AWS used without 'meaningful human control' raises serious legal, ethical and security concerns. The communiqué called for urgent negotiations on a legally binding instrument to regulate AWS.[19]

On 29–30 April Austria hosted a conference to discuss the legal and ethical challenges raised by AWS. More than 1000 representatives from 144 countries attended the conference, which concluded with a 'conference summary' that the Austrian government later submitted to the United Nations secretary-general's report on LAWS (see below).[20]

Report on LAWS by the UN secretary-general

In 2023 the UN General Assembly tasked UN Secretary-General António Guterres to seek the views of member states and observer states on LAWS (focusing on ways to address concerns raised by LAWS with respect to humanitarian, legal, security, technological and ethical perspectives)

[18] CCW Convention, GGE on LAWS, 'Rolling text, status date: 8 November 2024', 8 Nov. 2024.

[19] 'Communiqué of the Regional Conference on the Peace and Security Aspects of Autonomous Weapons Systems: An ECOWAS perspective', 17–18 Apr. 2024.

[20] 'Chair's summary', Humanity at the Crossroads: Autonomous Weapons Systems and the Challenge of Regulation Conference, Vienna, 30 Apr. 2024.

and to submit a report reflecting the full range of views.[21] The secretary-general received a total of 58 submissions from more than 73 countries, and another 33 submissions from the ICRC, civil society groups and the scientific community. The report was published in July 2024 and included views regarding definitions and characterizations; concerns and potential benefits; next steps; and observations and conclusions of the secretary-general, as well as an annex containing all of the submissions.[22] In his conclusion, Secretary-General Guterres stressed that 'time is running out for the international community to take preventative action on this issue', and he called for states to adopt a legally binding instrument to prohibit and regulate LAWS no later than the end of 2026.[23] Specifically, Guterres argued for the need to prohibit LAWS that function without human control or oversight, that cannot be used in compliance with IHL, and that are used to target humans (as the latter would be 'a moral line that must not be crossed'), and to regulate all other types of LAWS.[24] As for the way forward, the secretary-general encouraged the GGE on LAWS to 'work diligently to fulfil the mandate as soon as possible' and the General Assembly to continue discussions on LAWS.[25]

While the submissions to the secretary-general's report largely mirrored views already expressed at the CCW, the report was more comprehensive because it included views of states that are not parties to the CCW and expanded on concerns beyond IHL, being those related to ethics, human rights and international security. Thus, both the scope of the report and the secretary-general's significant recommendations were perceived by many states and civil society organizations as generating important momentum towards establishing a legally binding instrument regulating LAWS, whether within or outside the CCW.

General Assembly resolution on LAWS

In November 2024 the UN General Assembly First Committee adopted its second-ever resolution on LAWS.[26] The resolution, tabled by Austria and supported by a cross-regional group of 26 co-sponsoring states, recognized the serious humanitarian, legal, security, technical and ethical challenges that LAWS raise. A noteworthy element in the resolution is that it requests the UN secretary-general to facilitate informal consultations in 2025 to consider, among other things, the secretary-general's report on LAWS and the

[21] United Nations, General Assembly, 'Lethal autonomous weapons systems', A/C.1/78/L.56, 12 Oct. 2023.

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

[23] United Nations, General Assembly, A/79/88, para. 90.

[24] United Nations, General Assembly, A/79/88, para. 90.

[25] United Nations, General Assembly, A/79/88, para. 91.

[26] UN General Assembly Resolution 79/62, 2 Dec. 2024.

work of the GGE on LAWS.[27] The informal consultations, which are expected to take place over two days in New York, will be open to the participation of all UN member states and observer states, international and regional organizations, the ICRC and civil society, including the scientific community and industry. According to the resolution, the informal consultations are intended to supplement, not replace, discussions at the CCW. However, while several states continue to stress that the CCW remains the appropriate forum to address LAWS, the appeal of the General Assembly to many states is it allows the inclusion of a broader range of perspectives, spanning from ethical and human rights assessments to concerns around proliferation and impacts on global security and on regional and international stability.

Military AI beyond autonomous weapon systems

REAIM summit in Seoul

In 2024 the second international Summit on Responsible Artificial Intelligence in the Military Domain (REAIM 2024) was held in Seoul on 10–11 September. South Korea, Kenya, the Netherlands, Singapore and the United Kingdom co-hosted the summit.[28] The REAIM summit, initially launched in 2023 by the Netherlands, provides a platform for a wide range of stakeholders to discuss the governance of military AI. In 2024 the agenda was organized around three thematic streams: the impact of AI on international peace and security; implementing responsible AI in the military domain; and envisaging future governance of AI in the military domain.[29]

REAIM 2024 concluded with the adoption of a 'Blueprint for Action' for responsible AI in the military.[30] While 63 countries spanning all continents adopted this outcome document, some important states did not. Although participating in the 2024 summit, China, India and Israel decided not to join the call for action. Russia was not invited to the summit because of its full-scale invasion of Ukraine in 2022.[31] The Blueprint for Action is a soft-law document compiling 20 actions divided among the summit's three thematic streams. The actions aim to 'harness the benefits and opportunities of AI while adequately addressing the risks and challenges involved'.[32]

On the impact of AI on international peace and security—the first thematic stream—endorsing states recognized that military AI comes with benefits for

[27] UN General Assembly Resolution 79/62 (note 26), para. 7.

[28] REAIM Summit 2024, 'Overview', [n.d.].

[29] REAIM Summit 2024, 'Program', [n.d.].

[30] REAIM Summit 2024, 'Blueprint for Action', 11 Sep. 2024.

[31] Hong, W., 'US–China competition looms large at Seoul summit on use of AI in military', Asia Pacific foundation of Canada, 9 Oct. 2024.

[32] REAIM Summit 2024, 'Blueprint for Action' (note 30), paras 1–6.

military applications while acknowledging its potential risks.[33] The Blueprint for Action stresses that policy action should go beyond AI-enabled weapons and pay attention to 'AI-enabled decision support systems for combat operations, AI in cyber operations, AI in electronic warfare and AI for information operations', and stresses the 'need to prevent AI technology to be used to contribute to the proliferation of [weapons of mass destruction (WMDs)]'.[34]

The actions under the second thematic stream—implementing responsible AI in the military domain—reaffirm the validity of international law and other relevant legal frameworks in the research, design, development and implementation of military AI. This section also acknowledges some key principles for responsible AI in the military: namely, human responsibility and accountability, trustworthiness, appropriate human involvement, and the ability 'to understand, explain, trace and trust the outputs produced by AI capabilities in the military domain'.[35]

Finally, the actions under the third thematic stream—envisaging future governance of AI in the military domain—stress the importance of international cooperation and capacity building on responsible development, deployment and use of AI in the military domain. This section also acknowledges the existence of multiple initiatives that link to AI in the military domain and encourages a more 'synergistic and complementary' approach among them.[36]

Immediately after the summit, the Global Commission on Responsible Artificial Intelligence in the Military (GC REAIM) convened. Established after the first REAIM summit and composed of various experts in military AI, the GC REAIM aims to guide international discussions on responsible AI use in the military. Specifically, its goal is the promotion of 'mutual awareness and understanding among the many communities working on issues related to the global governance of AI in the military domain'.[37] Initially established for two years, this group will produce a 'strategic guidance report' that will cover: (*a*) what is meant by responsible AI in the military domain; (*b*) the elements that need to be included in the AI lifecycle to design, develop, produce, introduce and use AI in the military domain in a responsible way; and (*c*) the governance mechanisms that should be set up to design, develop and use AI in the military domain responsibly.[38] The GC REAIM will continue its work through several meetings planned in 2025.[39]

As of the end of December 2024, no states had announced a commitment to organize and host a third REAIM Summit.

[33] REAIM Summit 2024, 'Blueprint for Action' (note 30), paras 4 and 5.

[34] REAIM Summit 2024, 'Blueprint for Action' (note 30), p. 1.

[35] REAIM Summit 2024, 'Blueprint for Action' (note 30), para. 9.

[36] 'REAIM Summit 2024, 'Blueprint for Action' (note 30), para. 19.

[37] Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM), 'Mission statement', [n.d.].

[38] GC REAIM, 'Activities', [n.d.].

[39] GC REAIM, 'Conference timeline', [n.d.].

UN General Assembly resolution on military AI

The two original organizers of the REAIM summit, the Netherlands and South Korea, submitted in 2024 a joint resolution, co-sponsored by a cross-regional group of 21 states, to the UN General Assembly titled ‘AI in the military domain and its implications for international peace and security’.[40] The resolution was adopted in the First Committee on 24 December, with 159 in favour, 2 against (Russia and North Korea) and 5 abstentions (Belarus, Ethiopia, Iran, Nicaragua, Saudi Arabia).[41] (In opposing the resolution, Russia cited concerns about fragmenting multilateral processes like the GGE on LAWS, pre-empting future AI military regulations, unclear key terms, reliance on controversial criteria not in international law, and the resolution reflecting the views of a non-inclusive group of states.[42]) With this resolution and the REAIM summits, the Netherlands and South Korea have been seeking ‘to set the international standards for the military use of AI’.[43]

The resolution transposes the spirit of the Blueprint for Action to the UN context—with the notable difference that the resolution does not include reference to WMDs. The resolution’s text underlines the importance of adopting a multi-stakeholder approach for the governance of military AI and recognizes the crucial role of the private sector, civil society and academia in helping states and society understand how AI poses risks to peace and security. It reiterates the application of the Charter of the United Nations, IHL and international human rights law to military AI, and invites states to continue multilateral dialogues to address opportunities and challenges of AI’s integration in the military. The resolution also echoes the Global Digital Compact—a 2024 UN framework for the global governance of digital technologies and AI (see section III)—by inviting states to ‘bridge the divides between countries with regards to responsible AI in the military domain’ and calls on states to share good practices and lessons learned on the responsible application of AI in the military domain.[44]

In a similar vein as the 2023 resolution on LAWS, this resolution requests the UN secretary-general to collect states’ views on the opportunities and challenges that the integration of AI in the military domain poses to inter-

[40] UN General Assembly Resolution 79/239, 24 Dec. 2024.

[41] United Nations, ‘Fourteen new drafts, including on implications of artificial intelligence in the military domain, approved in First Committee by 34 votes’, Meetings coverage, GA/DIS/3757, 6 Nov. 2024.

[42] Representative of the Russian delegation to the United Nations, ‘Statement in explanation of vote on a draft resolution “Artificial intelligence in the military domain and its implications for international peace and security” L.43 in the First Committee of the 79th session of the UNGA’, Russian Ministry of Foreign Affairs, Foreign policy news, 6 Nov. 2024.

[43] Yeon Gyeong, Y., ‘UN committee adopts Korea-led proposal on military use of AI’, Korea.net, 8 Nov. 2024.

[44] UN General Assembly Resolution 79/239 (note 40), para. 6.

national peace and security, and to publish the views in a report.[45] However, here the focus is to explicitly be on areas *other* than LAWS. The UN secretary-general is also to seek input from international and regional organizations, the ICRC, civil society, the scientific community and industry, and 'to include these views in the original language received in the annex to the aforementioned report'.[46]

US political declaration

The US Political Declaration on the Responsible Military Use of Artificial Intelligence and Autonomy—launched in February 2023 at the REAIM summit in the Hague—establishes 10 foundational principles to foster responsible military AI.[47] Initially, the US political declaration received limited endorsement from other states, prompting revisions to broaden support.[48] One of the main changes was dropping the reference to AI in nuclear command and control, as some states considered this provision to be a legitimation of nuclear weapons and thus were reluctant to support the declaration.[49] A revised version of the political declaration was rolled out in November 2023 and at the end of December 2024, 58 states had endorsed it.[50]

In March 2024, the USA held the first plenary meeting of states endorsing the political declaration.[51] Endorsing states present at the meeting in Washington DC formed three working groups tasked with promoting the implementation of the political declaration through technical exchanges, sharing of best practices, and the development of technical standards.[52] Working group one on 'AI assurance' focuses on ensuring AI systems follow strict guidelines for defined uses, with rigorous testing, safeguards against errors, and human override capabilities.[53] The second working group on 'accountability' deals with the human aspect of AI governance: ensuring military personnel receive proper training to understand the technology's capabilities and limits, and access to clear, auditable documentation on its functionality.[54] The third

[45] UN General Assembly Resolution 79/239 (note 40), para. 7.

[46] UN General Assembly Resolution 79/239 (note 40), para. 8.

[47] US Department of State, Bureau of Arms Control, Deterrence and Stability, 'Political Declaration on the Responsible Military Use of Artificial Intelligence and Autonomy', [n.d.].

[48] Depp, M., 'The next step in military AI multilateralism', *Lawfare*, 26 Mar. 2024.

[49] Depp (note 48).

[50] US Department of State, Office of the Spokesperson, 'Undersecretary Jenkins rolls out the Political Declaration on Responsible Military Use of Artificial Intelligence (AI) and Autonomy', 13 Nov. 2023; and US Department of State, Bureau of Arms Control, Deterrence and Stability (note 47).

[51] US Department of State, Office of the Spokesperson, 'Inaugural plenary meeting of states endorsing the political declaration on responsible military use of artificial intelligence and autonomy', 19 Mar. 2024.

[52] United Nations, 'The Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy: Delivering concrete solutions', UN Web TV, 29 Oct. 2024.

[53] Freedberg, S. J., 'US joins Austria, Bahrain, Canada, & Portugal to co-lead global push for safer military AI', *Breaking Defense*, 28 Mar. 2024.

[54] United Nations (note 52), 00:12:02.

working group on 'oversight' focuses on broad policy issues, including mandating legal reviews to ensure compliance with IHL, establishing oversight by senior officials and eliminating unintended biases in AI systems.[55]

The US political declaration is framed as a complementary tool to other military AI governance initiatives and other specialized forums that address legal and ethical issues related to AI and AWS. With this declaration the US aims to align with like-minded states under shared principles and work towards their implementation.

III. Governing the challenges presented by civilian artificial intelligence

States have recognized that developments in civilian AI may negatively affect peace and security.[56] They intend to mitigate these risks across various forums. Notable multilateral efforts include UN-led processes on technology governance and the AI Safety Summit. This section highlights key developments in these two forums in 2024.

United Nations efforts towards global AI governance

The promises and perils of AI emerged as a crucial topic at the UN General Assembly in 2024. In his opening remarks to the 79th session, the UN secretary-general emphasized that AI technology brings both risks and benefits, positioning the UN as a key platform to ensure that AI is a force for good and to centralize international cooperation on AI's challenges.[57] During the general debate, leaders expressed concerns about existential risks stemming from AI such as the misuse of scientific advancements, the threat to democratic processes in the age of AI-enabled disinformation, and the inequalities stemming from 'AI use and development between developed and developing countries'.[58]

The 79th session of the General Assembly began with the Summit of the Future, which culminated with three outcome documents adopted under Resolution 79/1: the Pact for the Future, the Global Digital Compact and the Declaration on Future Generations.[59] The Pact for the Future features 56 actions to foster multilateralism which aim to address a broad range of

[55] Freedberg (note 53).

[56] British Government, 'The Bletchley Declaration by countries attending the AI Safety Summit', Policy paper, 1 Nov. 2023.

[57] United Nations Secretary-General, Address to the General Assembly, New York, 24 Sep. 2024, pp. 7, 9.

[58] Apt, C., 'AI at UNGA79: Recapping key themes', Just Security, 1 Oct. 2024.

[59] United Nations, *Pact for the Future, Global Digital Compact and Declaration on Future Generations*, Summit of the Future Outcome Documents, Sep. 2024; and UN General Assembly Resolution 79/1, 22 Sep. 2024.

issues, including peace and security, human rights, sustainable development and digital cooperation.[60] The Global Digital Compact is designed to bridge digital divides between developed and developing regions and to establish a foundation for global digital and AI governance.[61] In the section dedicated to peace and security in the Pact, action 27 is of relevance for AI governance. Among other pledges related to emerging technologies, states commit to developing an instrument on AWS (without prejudging its nature), enhancing cooperation to tackle digital divides, continuing to monitor the impact of military AI, and asking the UN secretary-general to keep reporting on the impacts of new and emerging technologies on peace and security.[62] This last commitment is related to objective 5 of the Global Digital Compact in which states agree to establish a multidisciplinary Independent International Scientific Panel on AI within the UN and to initiate a global dialogue on AI governance. The goal of this panel will be to 'promote scientific understanding through evidence-based impact, risk and opportunity assessments'.[63]

Both the Pact for the Future and the Global Digital Compact build on the recommendations of the UN secretary-general's High-Level Advisory Body on Artificial Intelligence (AIAB) report 'Governing AI for Humanity', which was presented at the beginning of the Summit for the Future.[64] The AIAB, an advisory group of 39 experts from all regions and multiple sectors, including government, the private sector and civil society, was established by the UN secretary-general in December 2023.[65] The AIAB report—which resulted from an extensive global consultation with over 2000 participants—makes key recommendations, including the establishment of a globally inclusive AI governance structure. The report also outlines seven steps to fill current regulatory gaps and urges cooperation across sectors to develop AI responsibly while protecting human rights.[66]

AI Safety Summit

In May 2024 South Korea hosted the Seoul AI Safety Summit—the second such summit, following the UK AI Safety Summit in November 2023.[67] The

[60] United Nations, Summit of the Future, 'Outcome document: Pact for the Future', [n.d.].

[61] United Nations, 'Global Digital Compact', [n.d.].

[62] United Nations, *Pact for the Future, Global Digital Compact and Declaration on Future Generations* (note 59), p. 19; and UN General Assembly Resolution 79/1 (note 59), para. 48.

[63] United Nations, *Pact for the Future, Global Digital Compact and Declaration on Future Generations* (note 59), p. 37; and UN General Assembly Resolution 79/1 (note 59), para. 56(a).

[64] United Nations, AI Advisory Body, *Governing AI for Humanity: Final Report* (United Nations: New York, Sep. 2024).

[65] United Nations, AI Advisory Body, 'About the UN Secretary-General's High-Level Advisory Body on AI', [n.d.].

[66] United Nations, AI Advisory Body, 'UN Secretary-General's High-Level Advisory Body on Artificial Intelligence releases proposals for global governance of AI', Press release, 19 Sep. 2024.

[67] British Government, 'AI Safety Summit 2023', [n.d.].

UK summit concluded with the adoption by many states, including China and the USA, of the Bletchley Declaration, in which countries recognize that the most advanced AI models (frontier models) could pose safety and security risks, particularly in the areas of cybersecurity, biotechnology and misinformation.[68]

The Seoul summit brought four key developments. First, the ministers of 27 states and the European Union (EU) jointly affirmed the Seoul ministerial statement for advancing AI safety, innovation and inclusivity.[69] This declaration broadens the scope of the AI Safety Summits to include innovation and inclusivity. Unlike the Bletchley Declaration, the Seoul ministerial statement was not affirmed by China, for reasons that are unclear, although some have speculated the summit could be viewed as 'promoting a Western-centric view of global AI governance'.[70]

Second, 10 states and the EU affirmed the Seoul Statement of Intent toward International Cooperation on AI Safety Science.[71] This statement of intent materialized in November 2024 when the USA held the first official meeting of the International Network of AI Safety Institutes.[72]

Third, 16 leading AI companies, including Anthropic, Google, Microsoft, Mistral AI, Naver, OpenAI and Zhipu.ai, agreed to the Frontier AI Safety Commitments.[73] These commitments—unlike other private sector commitments to advance AI safety, such as the Frontier Model Forum or the White House Commitments—involve companies from different geographical areas, including the USA and China.[74] The signatory companies promise to work—at the company level and voluntarily—on risk mitigation measures in the development of the most advanced models, and are expected to report on their progress at the next AI Safety Summit.[75]

The fourth key development of the Seoul Summit was the presentation of the interim version of the International Scientific Report on the Safety of Advanced AI, a report coordinated by Canadian computer scientist Yoshua

[68] British Government, 'The Bletchley Declaration by countries attending the AI Safety Summit' (note 56).

[69] British Department for Science, Innovation and Technology, 'Seoul ministerial statement for advancing AI safety, innovation and inclusivity: AI Seoul Summit 2024', Policy paper, 22 May. 2024.

[70] See e.g. Meltzer, J. P. and Triolo, P., 'The Bletchley Park process could be a building block for global cooperation on AI safety', Brookings research, 4 Oct. 2024.

[71] British Department for Science, Innovation and Technology, 'Seoul Statement of Intent toward International Cooperation on AI Safety Science, AI Seoul Summit 2024 (Annex)', Policy paper, 21 May 2024.

[72] O'Brien, M. and Ortutay, B., 'US gathers allies to talk AI safety as Trump's vow to undo Biden's AI policy overshadows their work', AP, 21 Nov. 2024.

[73] British Department for Science, Innovation and Technology, 'Frontier AI Safety Commitments, AI Seoul Summit 2024', Policy paper, 21 May 2024.

[74] Frontier Model Forum, 'Frontier Model Forum: Advancing frontier AI safety', [n.d.]; and White House, Briefing Room, 'Biden–Harris administration secures voluntary commitments from leading AI companies to manage the risks posed by AI', Statements and Releases: Fact Sheet, 21 July 2023.

[75] Elysée Palace, 'AI Action Summit—Presentation', [n.d.].

Bengio and involving a diverse group of AI experts from various countries.[76] This interim report summarizes current scientific knowledge on general-purpose AI and focuses on understanding and managing the risks that stem from it, which it divides into three categories: risks from malfunction, malicious risk and systemic risks.[77]

The next AI Safety Summit will be in Paris in February 2025—under the name 'Paris AI Action Summit'—where the intention is to concentrate on deliverable steps to implement the progress made at the first two summits.[78]

IV. Other important developments in the governance of artificial intelligence

Besides the two multilateral processes discussed in section III on the governance of the various risks that AI poses, several other important developments and initiatives also took place in 2024. Chief among these are the EU's efforts to regulate AI, the implementation of the USA's Executive Order on Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, and various Chinese AI governance initiatives.

European Union regulation of AI

In 2024 the EU adopted the Artificial Intelligence Act (AI Act)—the first binding regulation specifically about AI, thus marking a significant milestone in AI governance.[79] With this regulation, the EU aims to protect fundamental rights, democracy and the rule of law while promoting AI innovation in Europe.[80] The AI Act bans some applications of AI, including certain types of predictive policing, social scoring and AI that manipulates human behaviour. For other applications, it adopts a risk-based approach that sets different requirements for AI providers depending on the risks associated with their products.[81] The AI Act does not apply to products designed exclusively for military, defence and national security purposes.[82] However, the overlap between civilian and military AI development has led some experts to anticipate indirect effects on military AI innovation in Europe.[83] Another noteworthy element of the AI Act is its extraterritorial effect: it applies to all

[76] *International Scientific Report on the Safety of Advanced AI: Interim Report* (AI Seoul Summit 2024: London, May 2024).

[77] *International Scientific Report on the Safety of Advanced AI: Interim Report* (note 76), chapter 4.

[78] Elysée Palace (note 75).

[79] European Parliament, 'Artificial Intelligence Act: MEPs adopt landmark law', Press release, 13 Mar. 2024.

[80] European Commission, 'Artificial Intelligence—Questions and answers', 1 Aug. 2024.

[81] European Commission, 'AI Act', 14 Oct. 2024.

[82] European Commission, 'Artificial Intelligence—Questions and answers' (note 82).

[83] Greene, N., 'The EU AI Act could hurt military innovation in Europe', *Encompass*, Jan. 2024.

products deployed in the EU market or affecting an EU citizen, regardless of the provider's location. This extraterritorial aspect was included to influence international standards, in the same vein as the EU's precedent in shaping data protection norms.

To oversee the implementation of the AI Act, the European Commission established the European AI Office. Operational since June 2024, the AI Office assists AI governance bodies in EU member states, develops tools and methodologies to test general-purpose AI, promotes the development and use of 'trustworthy' AI, and fosters international cooperation.[84] Among other key actions in 2024, the AI Office initiated a multi-stakeholder consultation to develop guidelines for general-purpose AI models.[85] Part of this process focuses on elaborating practices to identify, assess and mitigate systemic risks. In the context of the AI Act, systemic risks refer to general-purpose AI's actual or potential negative impacts on security, safety, public health and fundamental rights. In 2025 the AI Office will continue its oversight efforts and is expected to release general-purpose AI guidance in the second half of the year.

Implementation of the US Executive Order on Safe, Secure and Trustworthy AI

On 30 October 2023 the White House released Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Throughout 2024, various US federal agencies began implementing its provisions. A major outcome of the executive order was the establishment of the US AI Safety Institute (US-AISI), hosted within the US National Institute for Standards and Technology (NIST). In February the US Secretary of Commerce announced the leadership of the US–AISI, and in May the institute released its strategy to advance the understanding and mitigation of AI risks.[86] The US-AISI's work has gained momentum over the year, by addressing issues like the safety of AI models used in chemical and biological research, engaging in international collaborations with other AI safety institutes, and setting up a task force to monitor the impact of AI on national security.[87]

[84] European Commission, 'European AI Office', 12 Dec. 2024.

[85] European Commission, 'AI Act: Have your say on trustworthy general-purpose AI', 18 Dec. 2024.

[86] US Department of Commerce, National Institute for Standards and Technology (NIST), 'US Commerce Secretary Gina Raimondo announces key executive leadership at US AI Safety Institute', News, 7 Feb. 2024; and US Artificial Intelligence Safety Institute (US-AISI), 'Strategic vision', 1 Oct. 2024.

[87] See e.g. 'Safety considerations for chemical and/or biological AI models', NIST Notice, *Federal Register*, 3 Dec. 2024; NIST, 'US AI Safety Institute and European AI Office hold technical dialogue', News, 12 July 2024; and British Department of Science, Innovation and Technology and US-AISI, 'UK & United States announce partnership on science of AI safety', Joint press release, 2 Apr. 2024.

The executive order also called for a National Security Memorandum (NSM) on AI, reflecting the US recognition that AI is a strategic technology for many countries.[88] The NSM prioritizes protecting US AI ecosystems, ensuring safe, secure and trustworthy AI, and reducing AI misuse globally. It also highlights the importance of private sector involvement in AI development and the need for government coordination. The NSM tasked the NIST and the US-AISI with being the primary links between the private sector and government on AI safety, particularly in areas like cybersecurity and biological and chemical weapons risks.[89]

These efforts rest on fragile legal grounds. The executive order could easily be overturned, and President-elect Trump vowed to repeal it, arguing that it stifles innovation and 'imposes Radical Leftwing ideas on the development of this technology'.[90] The future of AI governance under the new US administration remained unclear at the end of 2024, with some voices predicting a shift toward more Silicon Valley–friendly policies, fewer antitrust regulations and potentially greater fragmentation in AI governance.[91]

Chinese AI governance

The Chinese vision for AI governance was released in October 2023 with the adoption of the Global AI Initiative.[92] Since releasing this initiative and signing the Bletchley Declaration, China has advanced its national AI governance efforts by issuing the Basic Safety Requirements for Generative Artificial Intelligence Services in 2024.[93] The document proposes pre-deployment tests for a variety of safety and security risks across five categories, including threats to national security, discriminatory or false content, commercial violations, privacy breaches, and inaccuracies in critical fields like medicine and infrastructure.[94] According to AI safety researchers, Chinese researchers have also been active in developing AI safety evaluation tools, with a growing focus on frontier issues such as large language models (LLM) 'unlearning',

[88] White House, Briefing Room, 'Memorandum on advancing the United States' leadership on artificial intelligence; harnessing artificial intelligence to fulfill national security objectives; and fostering the safety, security, and trustworthiness of artificial intelligence', Presidential Actions, 24 Oct. 2024.

[89] White House, Briefing Room, 'Memorandum' (note 88).

[90] Trump 2024 Presidential Campaign, '2024 GOP Platform: Make America great again!', [n.d.], p. 9.

[91] Goudarzi, S., 'Trump's potential impact on emerging and disruptive technologies', *Bulletin of the Atomic Scientists*, 6 Nov. 2024; and Wiggers, K., 'What Trump's victory could mean for AI regulation', *Tech Crunch*, 6 Nov. 2024.

[92] Cyberspace Administration of China, 'Global AI governance initiative', 18 Oct. 2023.

[93] Center for Security and Emerging Technology, Translation of National Technical Committee 260 on Cybersecurity of Standardization Administration of China, 'Basic safety requirements for generative artificial intelligence services', Technical documentation, 4 Apr. 2024.

[94] Concordia AI, 'China's AI safety evaluation ecosystem', AI Safety in China Blog, 13 Sep. 2024.

misuse risks in biology and chemistry, and assessing 'power-seeking' and 'self-awareness' risks in LLMs.[95]

Another two noteworthy developments were the issue of a joint statement with France on AI governance and the first intergovernmental dialogue between the USA and China in May 2024.[96] In the joint statement, France and China acknowledge the opportunities and challenges presented by AI and commit to advancing the development of secure AI systems. They also commit to continue working within global AI governance efforts by supporting existing initiatives, such as the AIAB report 'Governing AI for Humanity' (see section III). Additionally, China expressed its willingness to participate in the Paris AI Action Summit in 2025.

In May 2024 a US and a Chinese delegation met in Geneva to discuss AI governance. China expressed support for stronger global governance of AI, emphasizing the UN's leading role and signalling a willingness to coordinate with the international community, including the USA, to establish globally accepted standards for AI governance.[97] However, China, in line with its traditional stance on these issues, also opposed US restrictions and pressures on its AI sector.[98] The USA, for its part, stressed the need for AI systems to be 'safe, secure and trustworthy' while highlighting concerns about the misuse of AI, specifically pointing to risks involving China, but without further elaboration.[99] In November 2024 President Joe Biden and President Xi Jinping met again in Peru. AI was on their agenda, and both leaders reiterated 'the need to maintain human control over the decision to use nuclear weapons'.[100]

V. Conclusions

AI advances are poised to bring enormous benefits, but they can also create or exacerbate existing threats to international peace and security. In recent years, states have increasingly acknowledged the need to manage these complex risks—stemming from both civilian and military AI—through the estab-

[95] Concordia AI (note 94).

[96] Elysée Palace, 'Déclaration conjointe entre la République française et la République de Chine sur l'intelligence artificielle et la gouvernance des enjeux globaux' [Joint declaration between France and China on artificial intelligence and the governance of global issues], 6 May 2024; and Keaten, J. and Chan, K., 'In first AI dialogue, US cites "misuse" of AI by China, Beijing protests Washington's restrictions', AP, 15 May 2024.

[97] Geopolitechs, 'China's readout of the first Sino–US intergovernmental dialogue on AI', 15 May 2024.

[98] Bromley, M., Mustafić, S. and Yuan, J., 'China takes aim at the export control regimes: Targeted critique or misguided attack?', *WorldECR*, no. 123 (Oct. 2023).

[99] Keaten and Chan (note 96).

[100] White House, Briefing Room, 'Readout of president Joe Biden's meeting with President Xi Jinping of the People's Republic of China', Statements and Releases, 16 Nov. 2024. On China–USA dialogue on nuclear weapons see chapter 8, section II, in this volume.

lishment of new forums and initiatives. In 2024 AI governance remained a key priority in international discussions, as states deepened their engagement with ongoing initiatives and solidified AI as a central topic for peace and security discussions. For example, both the second REAIM Summit and the second AI Safety Summit took place in 2024. In addition, in its 79th session, the UN General Assembly adopted key resolutions on military AI and LAWS. It is probable that 2024 will also be remembered as the year when important regional regulatory efforts like the EU AI Act were adopted. However, a big question flowing from the many developments relates to the extent to which various discussions, initiatives and resolutions will evolve as complementary or competing processes. While the outcomes of the various policy processes remain to be seen, what became clear in 2024—based on reports from several ongoing conflicts—is that the AI-related peace and security issues discussed in these various processes are more tangible than ever.