I. Governing the impact of artificial intelligence on international peace and security

VINCENT BOULANIN

For a decade, the international policy community discussed the impact of artificial intelligence (AI) on international peace and security almost exclusively through the prism of autonomous weapon systems (AWS).[1] This was because AWS remained the core technology of concern, even though AI enables the development of more advanced AWS. In 2023 that perspective changed. AI became a major concern its own right, as the international policy community came to realize that AI can impact international peace and security in many other ways beyond its role in the development of AWS.

This shift was reflected at the policy level by a series of developments that not only ended the monopoly of the 1981 Certain Conventional Weapons Convention (CCW Convention) on diplomatic talks on AI, but also gave the intergovernmental debate a broader and more high-level dimension.[2] These developments included the first-ever meeting of the United Nations Security Council on AI in July 2023 and the creation of two new discussion forums: the international summit on Responsible AI in the Military Domain (REAIM) and the AI Safety Summit.[3] REAIM provided a forum for states to discuss risks posed by the adoption of AI in the military domain beyond the sole case of AWS. In contrast, the AI Safety Summit provided a space for states to discuss the challenges that civilian AI—especially the most advanced forms thereof—could pose to peace and security. Both REAIM and the AI Safety Summit featured the participation of high-level decision makers—government ministers and in some cases heads of state—which was indicative of the political importance the topic gained in 2023.

This section retraces this evolution in greater detail. After taking stock of the policy developments specifically related to AWS, it discusses in chrono-

[1] 'International policy community' is used here as a shorthand to refer collectively to the governmental and non-governmental experts—e.g. from civil society and academia—who contribute to the debate on international policy matter in international forums such as those under the auspices of the United Nations or regional organizations like the European Union. On earlier discussions on the regulation of AWS see *SIPRI Yearbook 2014*, pp. 423–31; *SIPRI Yearbook 2017*, pp. 559–61; *SIPRI Yearbook 2018*, pp. 383–86; *SIPRI Yearbook 2019*, pp. 449–57; *SIPRI Yearbook 2020*, pp. 502–12; *SIPRI Yearbook 2021*, pp. 518–24; *SIPRI Yearbook 2022*, pp. 532–44; and *SIPRI Yearbook 2023*, pp. 463–70.

[2] 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), including lists of the states parties that have ratified the original, amended and additional protocols, see annex A, section I, in this volume. On other developments in the CCW Convention in 2023 see chapter 10, section I, in this volume.

[3] United Nations, 'International community must urgently confront new reality of generative artificial intelligence, speakers stress as Security Council debates risks, rewards', Meetings coverage, 9381st meeting (AM), SC/15359, 18 July 2023; Dutch Government, 'About REAIM 2023', [n.d.]; and AI Safety Summit website.

logical order the REAIM Summit, the first session of the Security Council on AI, and the AI Safety Summit. The section concludes by discussing what these events mean for the international governance of AI.

Governing the challenges presented by autonomous weapon systems

In 2023 the CCW Convention remained the epicentre of the intergovernmental conversation on AWS, but it lost its monopoly when the debate on AWS was extended beyond the CCW discussions in Geneva. Costa Rica, Trinidad and Tobago, and the Philippines initiated discussions on AWS at the regional level through conferences targeted at countries from Latin America, the Caribbean and the Indo-Pacific.[4] More importantly, the First Committee (Disarmament and International Security) of the UN General Assembly adopted a resolution that made AWS a formal discussion item for future meetings of the General Assembly in New York.

The 2023 meetings of the group of governmental experts at the CCW

The intergovernmental debate on AWS at the CCW marked its 10th anniversary in 2023. Since 2017 a group of governmental experts (GGE) has led the discussions. Under chair Ambassador Flávio Soares Damico from Brazil, the 2023 meeting of the GGE was due to explore further concrete proposals for developing the 'normative and operational framework governing AWS'.[5] In particular, the chair expected the GGE to consider the value and feasibility of a two-tiered regulatory approach. Such an approach would prohibit certain types of AWS, on the one hand, and place specific limits and requirements on the development and use of all other types of AWS, on the other. The GGE had already discussed that approach in 2022 but had failed to find sufficient common ground on the topic.[6]

The group met for two sessions of five days, on 6–10 March and 15–19 May 2023.[7] The first session was dedicated to state proposals and measures, with the first day reserved for general statements, and the rest of the week

[4] Conferencia Latinoamericana y del caribe sobre el impacto social y humanitari de las armas autónomas [Latin American and Caribbean Conference on the social and humanitarian impact of autonomous weapons], la Ribera de Belén, Costa Rica, 23–24 Feb. 2023; Caricom Conference 2023: The Human Impact of Autonomous Weapons, Port of Spain, Trininad and Tobago, 5–6 Sep. 2023; and Philippine Department of Foreign Affairs, 'Philippines calls for Indo-Pacific voices to address lethal autonomous weapons systems risk', Press release, 15 Dec. 2023.

[5] CCW Convention, Group of governmental experts on emerging technologies in the area of lethal autonomous weapons systems (GGE on LAWS), Report of the 2023 session, CCW/GGE.1/2023/2, 24 May 2023, para. 1.

[6] Boulanin, V., 'Intergovernmental efforts to address the challenges posed by autonomous weapon systems', *SIPRI Yearbook 2023,* pp. 467–70; and Acheson, R., 'Denial cannot stop the reality of momentum', *CCW Report*, vol. 10, no. 9 (28 July 2022).

[7] CCW Convention, CCW/GGE.1/2023/2 (note 5), para. 3.

was organized along (the usual) topical lines.[8] It was difficult for long-time observers of the GGE not to experience a strong sense of déjà vu. For each topic area, the GGE reiterated past disputes: (*a*) whether the GGE should aim for a legally binding instrument or voluntary measures; (*b*) whether a definition of AWS is a prerequisite for starting negotiation on a regulation; (*c*) whether new regulatory measures should only be guided by international humanitarian law (IHL) or also be informed by ethical considerations not already captured in law; (*d*) whether human control should be a pivotal concept; and (*e*) whether further guidance is needed on how to conduct legal reviews and risk mitigation measures, and whether these should be left to the discretion of states.

Despite these enduring differences, a sense of progress was palpable, at least in terms of political engagement. Most states showed clear support for the two-tiered approach, and most of the written proposals tabled ahead of the meeting were based on that approach, typically outlining elements that were deemed to be explicitly prohibited as well as elements that articulated positive and negative obligations for development and use of all AWS. Most states also signalled a willingness to engage in a constructive spirit and look for compromise language that would allow the GGE to have a tangible outcome.

Consideration of the chair's draft report

The richness of the proposals and exchanges allowed the GGE chair to return before the GGE's second session in May with an ambitious draft report which attempted to articulate compromise on three elements: (*a*) the characterization of AWS; (*b*) the basis on which AWS may be deemed off-limits; and (*c*) measures required to ensure compliance with IHL, ethical considerations and accountability, and to mitigate the risks of unintended engagement.[9]

The chair's draft report characterized AWS as weapon systems 'that, once activated, are able to identify, select, track and apply force to targets without further intervention'.[10] The proposed language was intended to accommodate those states that wanted conceptual clarity, such as Türkiye and India.[11] Although the wording reflected the most frequently used working definition

[8] Item 5 on the agenda: Topic 1 Characterization of LAWS—definitions and scope; Topic 2: Application of International Humanitarian Law (IHL): Possible prohibitions and regulations; Topic 3: Human–machine interaction/meaningful human control/human judgments and ethical considerations; Topic 4: Responsibility and accountability; Topic 5: Legal reviews; Topic 6: Risk mitigation and confidence-building measures; Topic 7: Any other subjects raised by delegations. CCW Convention, GGE on LAWS, Indicative timetable for the first session of the group, 22 Feb. 2023; and CCW Convention, CCW/GGE.1/2023/2 (note 5), paras 13–14.

[9] CCW Convention, GGE on LAWS, Draft report of the 2023 session, CCW/GGE.1/2023/CRP.2, 6 May 2023.

[10] CCW Convention, CCW/GGE.1/2023/CRP.2 (note 9), para. 19.

[11] CCW Convention, GGE on LAWS, Third meeting, First session, 7 March 2023, UN Web TV, 00:10:50–00:11:50 (Türkiye) and 01:27:00–01:29:00 (India).

of AWS, it did not find consensus. China and the Russian Federation opposed it. For China, the language was not sufficiently clear, while for Russia the group had not sufficiently discussed the characterization of AWS to include such wording in the report.[12]

Regarding the basis on which AWS may be prohibited, the chair suggested language that was focused on IHL considerations: 'Such weapons systems must not be deployed or used if their effect in attacks cannot be anticipated and controlled, as required by international humanitarian law in the circumstances of their use.'[13]

The absence of any reference to ethics and human control—due to insufficient support—was criticized by many states, including Austria, Cuba, Denmark, Ecuador, Germany, Guatemala, Palestine, Pakistan, Panama, Peru, Philippines, Sri Lanka and Sweden.[14] The rest of the discussion on the paragraph focused on the inclusion of the words 'anticipated and controlled'. Some states, notably Ecuador, Austria and Cuba, wanted further elaboration and proposed the following wording: 'weapons that cannot be predictable, reliable, understandable, explainable and traceable should be prohibited'.[15] The United States opposed this addition on the grounds that these terms are not found in existing IHL requirements, while Russia proposed deleting the sentence altogether.[16] The GGE eventually agreed on a redacted, rather tautological formulation, which in substance asserts that states should not use inherently unlawful weapons: 'Weapons systems based on emerging technologies in the area of LAWS must not be used if they are incapable of being used in compliance with IHL.'[17]

The part of the draft dedicated to the identification of limits and requirements for the development and use of AWS was extensive. It included several paragraphs and sub-paragraphs that prescribed concrete limits and requirements on the design and use of AWS to ensure compliance with IHL and accountability, as well as to mitigate the risk of unintended engagements. These could give the impression that the chair saw the second tier as the most important one from a normative standpoint. The draft language revolved around IHL but included elements that tended, indirectly, to satisfy ethical considerations.

In that regard, it was notable that the draft text referred to some of the ethical principles for the responsible design of AI, such as reliability, explainability and traceability: 'Ensure that the system is sufficiently *predictable*, *reliable*,

[12] CCW Convention, GGE on LAWS, Third meeting, Second session, 16 May 2023, UN Web TV, 02:30:00–02:38:00.
[13] CCW Convention, CCW/GGE.1/2023/CRP.2 (note 9), para. 20.
[14] CCW Convention, GGE on LAWS, third Meeting, Second session (note 12), 02:00:00–02:02:00.
[15] CCW Convention, GGE on LAWS, Third meeting, Second session (note 12), 02:05:00–02:07:00.
[16] CCW Convention, GGE on LAWS, Third meeting, Second session (note 12), 02:30:00–02:38:00.
[17] CCW Convention, CCW/GGE.1/2023/2 (note 5), para. 21(b).

understandable and *explainable*, and *traceable* by assessing how the weapon system is expected to perform in the anticipated circumstances of its use.'[18]

These concepts received a lot of attention. Interestingly, the states that pushed back the most on the inclusion of these elements were those that have already adopted such principles at the national level, such as the USA and the United Kingdom.[19] At the multilateral level, however, both states considered that these principles should not be included because they were not grounded in IHL.[20] Russia and Israel pushed for the deletion of these references for the same reasons. States that wanted this wording to be retained, notably Austria and Mexico, pointed out that IHL should not be invoked to limit the adoption of new requirements and that the very purpose of the CCW is to continue the codification of IHL.[21]

Similar exchanges occurred when the GGE discussed the language that specifically addresses the requirements for human–machine interaction. The chair's original proposal was that 'Those responsible for the use of a weapons system based on emerging technologies in the area of LAWS must be in a position, where feasible, to interrupt, disable or otherwise control the system or system functions, as necessary to comply with international humanitarian law.'[22]

The USA and the UK argued that IHL does not require persons to be 'in a position to interrupt' and that it sets a higher standard than currently exists for other weapon systems. They consequently proposed language that was more aligned with existing IHL language, notably the principle of precautions in attack.[23] Eventually, after several revisions and informal consultations, states agreed on a more abstract formulation which merely affirmed that 'Control . . . is needed to uphold compliance with international law, in particular IHL, including the principles and requirements of distinction, proportionality and precautions in attack.'[24]

[18] CCW Convention, CCW/GGE.1/2023/CRP.2 (note 9), para. 21(d) (emphasis added).

[19] The US Department of Defense (DOD) was the first to adopt ethical principles for responsible use of AI in the military domain. These principles include responsibility, equitability, traceability, realiability and governability. See US DOD, 'DOD adopts ethical principles for artificial intelligence', Press release, 24 Feb. 2020. The UK adopted a Defense AI Strategy that set out some principles for 'ambitious, safe and responsible' adoption of AI by its ministry of defence and armed forces. See UK Ministry of Defence, 'Ambitious, safe, responsible: Our approach to the delivery of AI-enabled capability in Defence', Policy paper, 15 June 2022.

[20] CCW Convention, GGE on LAWS, Fourth meeting, Second session, 16 May 2023, UN Web TV, 02:12:00–02:13:00.

[21] CCW Convention, GGE on LAWS, Fourth meeting, Second session (note 20), 00:40:00–00:42:00.

[22] CCW Convention, CCW/GGE.1/2023/CPR.2 (note 9), para. 23.

[23] CCW Convention, GGE on LAWS, Fourth meeting, Second session (note 20).

[24] CCW Convention, CCW/GGE.1/2023/2 (note 5), para. 21(c).

Some states, like Austria, deplored that the qualifier 'human' before the word 'control' was removed, but it was the compromise states had to make to keep an explicit reference to the notion of control in the list of requirements.[25]

The language suggested by the chair on accountability and risk mitigation had a more dramatic fate. Regarding accountability, the draft declared that states should ensure, among other things, operation of AWS 'within a responsible chain of command and control'; system design 'so as to ensure the attribution of responsibility'; 'adequate training to users'; and 'mechanisms for reporting and investigating incidents'.[26] The proposed wording on mitigating the risk of unintended engagement was that systems should be 'sufficiently predictable and reliable, allowing for an understanding of the expected consequences' of use.[27] None of this language survived the discussions, except for the reference to the 'training' of human operators. Russia was one of the states that opposed the inclusion of these elements on the basis that they are prerogatives of states and there was no need to standardize them.[28] India was also critical of the language that touched on the attribution of responsibility, and opposed the proposed mechanism to investigate incidents.[29]

Eventually, states agreed on a set of substantial conclusions that was much slimmer than originally intended but that nonetheless reflected the spirit of the two-tiered approach. On the final day of the second session, the GGE had to agree on a new mandate to move forward.[30] As in previous years, the dividing question was whether the language of the current mandate was sufficient or whether it was time for states to adopt a more ambitious formulation. The USA was of the view that the possibilities offered by the current mandate had not been exhausted and that the mandate was flexible enough to accommodate different positions on the desired policy outcome. Others, notably Austria, Cuba, Panama and the Philippines, wanted the new mandate to mark the beginning of a negotiation towards a legally binding instrument. The chair, in the spirit of compromise, had recommended that the group 'develop a single set of measures', but that formulation was questioned by the USA and a few other states.[31]

The GGE did not manage to come to a consensus on what the new mandate should be, so the discussion was concluded at the Meeting of High Contracting Parties to the CCW in November. States eventually agreed

[25] CCW Convention, GGE on LAWS, Eighth meeting, Second session, 18 May 2023, UN Web TV, 00:30:50–00:31:14.

[26] CCW Convention, CCW/GGE.1/2023/CPR.2 (note 9), para. 24.

[27] CCW Convention, CCW/GGE.1/2023/CPR.2 (note 9), para. 26.

[28] CCW Convention, GGE on LAWS, Sixth meeting, Second session, 17 May 2023, UN Web TV, 01:44:00–01:47:00.

[29] CCW Convention, GGE on LAWS, Sixth meeting, Second session (note 28), 00:02:40–00:03:20.

[30] CCW Convention, GGE on LAWS, 10th meeting, Second session, 19 May 2023, UN Web TV.

[31] CCW Convention, CCW/GGE.1/2023/CPR.2 (note 9), para. 30(a); and CCW Convention, GGE on LAWS, 10th meeting, second session (note 30), 02:46:00–02:47:00.

on compromise language which gives the GGE the mandate to consider and formulate, by consensus, measures to address AWS, 'including a set of elements of an instrument, without prejudging its nature'.[32] The meeting gave the GGE 20 days over two years to complete its work and expects the group to submit its report in time for the Seventh Review Conference of the CCW, which is scheduled for December 2025.

From Geneva to New York: Preventing the CCW becoming a single point of failure

For states that wanted to see more tangible progress towards a legally binding instrument on AWS, the outcome of the GGE was disappointing. Several states have been warning for years that if the GGE continued to underdeliver, they would take the policy conversation into a different forum. In 2023 that step was officially taken.

A group of states, under the leadership of Austria, tabled a resolution on AWS at the 78th session of the First Committee of the UN General Assembly in October 2023.[33] The resolution aimed to express the wide set of concerns associated with AWS—not just humanitarian concerns—and to formally put the topic on the agenda of future sessions of the General Assembly.[34] As an initial deliverable, the resolution requested the UN secretary-general to seek the views of member states and produce a report ahead of the 79th session of the General Assembly in October 2024.[35]

Member states overwhelmingly approved the resolution on 1 November 2023: 164 voted in favour; 5 voted against (Belarus, India, Mali, Niger and Russia); and 8 abstained (China, Iran, Israel, North Korea, Saudi Arabia, Syria, Türkiye and United Arab Emirates).[36] The list of those states that opposed or abstained from the vote partly overlaps with the list of CCW states parties that opposed the adoption of new regulatory measures on AWS. It was notable, in that regard, that military powers like the USA, the UK and France supported the resolution, even though it was conditional on the GGE being referenced as the main forum for discussing AWS. Their approval can be interpreted as an admission that the stakes are too high for letting the

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

[33] United Nations, General Assembly, First Committee, 78th session, Agenda item 99, 'Lethal autonomous weapons systems', Draft resolution by Austria et al., A/C1./78/L.56, 12 Oct. 2023 .

[34] The resolution expresses concern 'about the possible negative consequences and impact of autonomous weapon systems on global security and regional and international stability, including the risk of an emerging arms race, lowering the threshold for conflict and proliferation, including to non-State actors'. United Nations, General Assembly, A/C1./78/L.56 (note 33), p. 1.

[35] United Nations, General Assembly, A/C1./78/L.56 (note 33), para. 2.

[36] United Nations, General Assembly, First Committee, 78th session, 'First Committee approves new resolution on lethal autonomous weapons, as speaker warns "an algorithm must not be in full control of decision involving killing" ', Meetings coverage, 28th meeting (AM), GA/DIS/3731, 1 Nov. 2023.

CCW become a single point of failure for the regulatory future of AWS.[37] The creation of a discussion track within the General Assembly not only adds some redundancy to the existing intergovernmental efforts to regulate AWS, but also provides time and space for states to discuss challenges that would not squarely fit into the mandate of the CCW, such as issues about strategic stability and proliferation.

It is too early to say whether the General Assembly discussions may lead to a specific regulation on AWS. The resolution only marks the beginning of a formal conversation that seems likely to be hampered by the same conceptual and political disputes as the CCW process. The road will therefore be difficult, but the fact that the General Assembly is not required to operate by consensus may allow states the space to work towards the adoption of more substantial outcomes than under the CCW.[38]

The Summit on Responsible Artificial Intelligence in the Military Domain

The Summit on Responsible Artificial Intelligence in the Military Domain (REAIM) was the first international summit entirely dedicated to the military use of AI. The summit, which was held in the Hague on 15–16 February 2023, was co-hosted by the Netherlands and the Republic of Korea (South Korea).

The summit stemmed from a decision of the Dutch Parliament in 2021 which mandated that the government play a leading role in setting global norms for the responsible use of AI in the military domain.[39] The Dutch government consequently decided to organize a major event which would serve as a 'first step in a global dialogue' on the topic.[40] The event would complement, rather than replicate, the expert discussions on AWS within the CCW.[41] The intention was to enable a conversation on the benefits and risks associated with the full spectrum of applications of AI in the military domain, not just the use of AI in AWS. These other applications include the use of

[37] Single point of failure is a concept in engineering that refers to a part of a system that, if it fails, will stop the entire system from working.

[38] Unlike the CCW, which works on a consensus basis, the General Assembly can operate via a two-thirds majority voting rule. Also, while the CCW only includes states parties to the convention, the General Assembly includes the UN's entire membership.

[39] Quell, M., 'Countries sign military AI pact at historic summit, but is it enough?', *Courthouse News Service*, 16 Feb. 2023.

[40] De Han, T., REAIM Programme Coordinator at the Dutch Ministry of Foreign Affairs, quoted in de Boer, Y. 'Preparing for the military deployment of AI', *NOW*, 7 Feb. 2023.

[41] Sterling, T. and van den Berg, S., 'Dutch host first summit on "responsible" use of AI in the military', Reuters, 14 Feb. 2023.

AI for command and control, logistics and maintenance, personnel management, protection of civilians, and search and rescue.[42]

The summit largely fulfilled the objectives set by the organizers, namely to: (*a*) 'put the topic of responsible AI in the military domain higher on the political agenda'; (*b*) mobilize and activate a wide group of stakeholders to contribute to concrete next steps; and (*c*) foster and increase knowledge by sharing experiences, best practices and solutions.[43]

The summit succeeded in being a high-profile and agenda-setting event, with the participation of representatives from 80 governments, some at ministerial level. It culminated with the adoption of a 'call for action' to set the basis for an international dialogue on the responsible use of AI in the military domain. The key element of the call, which 57 of the 80 participating countries endorsed, was to establish a global commission on AI 'to raise all-round awareness, clarify how to define AI in the military domain and determine how this technology can be developed, manufactured and deployed responsibly'.[44] The USA also used the occasion to present a political declaration on the responsible military use of AI and autonomy that it had worked on with a few other countries.[45]

With the participation of 2500 representatives from government, the private sector, academia and civil society, the summit was also much larger and more diverse than any of the established UN-led meetings and events that touch on the governance of AI.[46] For example, the meeting of the GGE usually involves around 200 participants, mostly from governments.

The format of the summit was also conducive to information sharing and solution-oriented discussions. The programme was dense and very different from traditional arms control meetings. It featured plenary sessions, breakout sessions, an online talk show, an academic forum, a wargaming exercise and product demonstrations. These different elements covered a wide range of themes, including myth-busting AI, responsible deployment and use of AI, and governance issues.[47]

While the REAIM summit appeared to meet the objectives of the Dutch government, three criticisms of the event are notable.[48] First, some partici-

[42] See e.g. Saltini, A., *AI and Nuclear Command, Control and Communications: P5 Perspectives* (European Leadership Network: London, Nov. 2023); and Grand-Clément, S., *Artfificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain* (UNIDIR: Geneva, Oct. 2023).

[43] Dutch Ministry of Foreign Affairs, 'About REAIM 2023', [n.d.].

[44] Dutch Government, 'Call to action on responsible use of AI in the military domain', News item, 16 Feb. 2023.

[45] 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.].

[46] Dutch Government, 'Call to action on responsible use of AI in the military domain' (note 44); and REAIM Summit, Twitter, 17 Feb. 2023, <https://x.com/REAIMsummit/status/1626497947914407936?s=20>.

[47] Dutch Ministry of Foreign Affairs (note 43).

[48] Author's personal observations while a participant at the summit.

pants, mainly representatives of civil society, were dissatisfied with the framing of the event, which was in their view too focused on the military benefits of AI. Some argued, for example, that arms-producing companies displaying their products sent the wrong signal and legitimized the military use of AI. Others noted that the 'call for action' talked about the military use of AI in a highly positive way and avoided addressing the topic of AWS.[49]

Second, some state representatives complained that the drafting process for the 'call for action' was not inclusive, as it involved only a limited group of mainly Western states. The final text was also circulated very late, which prevented some states from properly studying the text. Compounding this issue was the fact that states were listed as endorsers of the call unless they had actively opted out. India, Brazil and South Africa were among the states that decided not to endorse the call.

Third, several observers judged the content of the call to be weak, as it did not go beyond what states had already agreed with the 11 guiding principles on AWS that the CCW adopted in 2019.[50] The Campaign to Stop Killers Robots, a coalition of non-governmental organizations that seeks to pre-emptively ban AWS, pointed out that the 'call for action' not only has no legally binding status but also represents a missed opportunity to initiate a serious discussion about new rules and limits.[51]

South Korea, which will host the next REAIM summit in 2024, appeared to take notice of some of the criticism. While it has yet to present the agenda and format for the 2024 summit, South Korea announced in December 2023 that it will conduct some regional consultations during 2024 with the aim of ensuring a more inclusive process.[52] The outcome of the next REAIM summit will likely depend on what emerges from these consultations. An open question in this context is whether South Korea will seek to use the political declaration presented by the USA at the first summit as a basis for a more substantial outcome. The role of the proposed global commission on AI remains unclear as of December 2023: neither its composition nor its programme of work had been announced.

Only time will tell whether and how the REAIM summit develops a role in international governance of AI, but the inaugural event succeeded in initiating a global conversation on military use of AI beyond the case of AWS.

[49] Qiao-Franco, G., 'Has REAIM "re-aimed" AI applications in the miltiary domain?', AutoNorms, 24 Feb. 2023.

[50] Qiao-Franco (note 49); and CCW Convention, GGE on LAWS, Report of the 2019 session, CCW/GGE.1/2019/3, 25 Sep. 2019, annex IV.

[51] DeGeurin, M., 'The first-ever international killer robots summit was human rights flop', *Gizmodo*, 18 Feb. 2023.

[52] Announcement made at the 22nd Republic of Korea–United Nations Joint Conference on Disarmament and Non-Proliferation, which focused on military AI. United Nations, 'UN, Republic of Korea host twenty-second Joint Conference on Disarmament and Non-Proliferation, focusing on military artificial intelligence', Press release no. DC/3865, 4 Dec. 2023.

Meeting of the UN Security Council on artificial intelligence

On 18 July 2023 the UN Security Council, on the initiative of the UK (which had the presidency at the time), held its first-ever formal session on AI and international peace and security. The inclusion of AI on the agenda was partly the result of two related events that occurred within a few months of each other.

First, in November 2022 the US-based organization OpenAI released a free preview of ChatGPT, an AI chatbot, that received widespread media coverage and millions of subscribers. That success and the mushrooming of other generative AI tools significantly raised awareness about AI developments among policymakers and the public.[53]

Second, in March 2023 the Future of Life Institute presented an open letter, which was signed by prominent figures in the AI community, that called for a six-month pause on 'giant' AI experiments.[54] The letter warned that current and future advances in AI could 'pose profound risks to society and humanity'. The institute also noted in its policy recommendations that generative AI tools could be misused to facilitate 'the inexpensive development of chemical, biological and cyber weapons'.[55] This proposed pause turned out to be controversial in the AI expert community. Some questioned the premise, notably the emphasis on the existential risks presented by 'advanced' AI—or 'artificial general intelligence' (AGI; a contested term that refers to general-purpose AI systems that outperform humans at a wide range of tasks)—of which some argue that ChatGPT is an early form.[56] Others expressed scepticism about the feasibility and effectiveness of the pause.[57] Nonetheless, the proposal succeeded in putting the conversation about AI risk in the spotlight. It triggered a wave of opinion columns and expert interviews in mainstream media worldwide about the danger AI poses and the need for regulation of the AI industry.[58] The public discussion also triggered reactions in the policy community up to the level of the UN.

The discussion at the UN Security Council on 18 July 2023 was focused on generative AI and how AI tools like ChatGPT could be misused for malicious

[53] On potential battlefield transformations from generative AI see Feldstein, S., 'The consequences of generative AI for democracy, governance and war', *Survival*, vol. 65, no. 5 (2023).

[54] Future of Life Institute, 'Pause giant AI experiments: On open letter', 22 Mar. 2023.

[55] Future of Life Institute, 'Policymaking in the pause: What can policymakers do *now* to combat risks from advanced AI systems?', Apr. 2023, p. 4.

[56] See thread by Bender, E., Twitter, 29 Mar. 2023, <https://x.com/emilymbender/status/1640920936600997889?s=20>. See also Rogers, R., 'What's AGI, and why are AI experts skeptical?', *Wired*, 20 Apr. 2023.

[57] See DeepLearning.AI, 'Yann LeCun and Andrew Ng: Why the 6-month AI pause is a bad idea', YouTube, 7 Apr. 2023.

[58] See e.g. Marcus, G. and Reuel, A., 'The world needs an international agency for artificial intelligence, say two AI experts', *The Economist*, 18 Apr. 2023; and *60 Minutes*, '"Godfather of AI" Geoffrey Hinton: The 60 Minutes interview', YouTube, 9 Oct. 2023.

political purposes. António Guterres, the UN secretary-general, noted in his remarks at the meeting that generative AI's advent 'could be a defining moment for disinformation and hate speech', and that its structural impact could aggravate existing 'social, digital and economic divides'.[59] One of the invited experts, Jack Clark, co-founder of the AI safety and research company Anthropic, drew attention to the fact that because generative AI is largely dual use, it could also be misused to develop weapon systems, including weapons of mass destruction like biological weapons.[60]

The discussion also touched on the risks associated with the military use of AI, notably the risk that the development or adoption of certain capabilities, such as AI in nuclear command and control, could disrupt global stability.

Not all members of the Security Council welcomed the discussion. Russia questioned the organ's relevance for discussing AI risks and pointed out that the CCW was the appropriate forum in which to discuss the military use of AI.[61]

The question of whether the AI community, particularly commercial organizations, can be trusted to develop AI responsibly was also a major discussion theme. The AI expert from the private sector, Jack Clark, was of the view that the development of AI could not be left solely to the private sector, and that states had to come together to regulate and control its development.[62] Several states echoed this view by pointing in their interventions to the regulatory efforts being undertaken at the national and regional levels. France, for instance, mentioned the discussion in the European Union (EU) around the adoption of an EU Artificial Intelligence Act which aims to put some guardrails on the development and deployment of AI systems.[63]

The UN secretary-general, in his speech, also announced the creation of a high-level advisory body on AI to explore possible options for the regulation of AI at the international level.[64] The group was formed in October 2023 and is due to present its findings and recommendations in Summer 2024.[65] The UK, which chaired the meeting, also announced in June 2023 that it would hold the first major summit on AI safety.[66]

[59] United Nations, SC/15359 (note 3).

[60] United Nations, SC/15359 (note 3); and UN Security Council, 'Artificial intelligence: Opportunities and risk for international peace and security—Security Council, 9381st meeting', 18 July 2023, UN Web TV, 00:16:00–00:17:00.

[61] UN Security Council (note 60), 02:00:00–02:01:00.

[62] UN Security Council (note 60), 00:18:00–00:22:00.

[63] UN Security Council (note 60), 01:35:00–01:36:00 . The EU's draft Artificial Intelligence Act was agreed in December 2023. See Council of the EU, 'Artificial Intelligence Act: Council and Parliament strike a deal on the first rules for AI in the world', Press release, 9 Dec. 2023.

[64] United Nations, 'Secretary-general urges Security Council to ensure transparency, accountability, oversight in first debate on artificial intelligence', Press release no. SG/SM/21880, 18 July 2023.

[65] United Nations, Office of the Secretary-General's Envoy on Technology, 'High-level Advisory Body on Artificial Intelligence', [n.d.].

[66] British Government, Office of the Prime Minister, 'UK to host first global summit on artificial intelligence', Press release, 7 June 2023.

The AI Safety Summit

The AI Safety Summit was held at Bletchley Park in the UK on 1–2 November 2023. Prime Minister Rishi Sunak's announcement of the summit was attributed to his desire to showcase the UK as an important player in the AI domain.[67] The summit was noteworthy in at least five respects.

First, it had a rather narrow thematic scope, focusing on the opportunities and risks associated with 'frontier AI systems'. These were defined as 'highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models'.[68]

The summit also dealt with only civilian applications of such systems and did not touch on military end-use. This niche focus was interpreted as an effort to avoid repeating or competing with existing processes, not least the general discussion on AI regulation in the EU, but also the debates on the military use of AI in the CCW and the UN Security Council. The summit was also viewed as a reflection of the policy community's increasing interest in the debate on existential risks presented by AGI to which 'frontier' AI systems may lead.

Second, the summit was relatively exclusive, bringing together 'leading AI nations, technology companies, researchers and civil society groups'.[69] That decision was arguably motivated by the fact that only a limited number of actors have the capabilities to develop frontier AI systems. The summit was further divided into two segments: multi-stakeholder expert discussion on the first day and a high-level discussion featuring only senior representatives from government and the private sector on the second day.

Third, the summit's main objectives were technical and operational in nature. The purpose was not to discuss possible regulatory guardrails but rather to agree on a 'ground-breaking plan for AI safety testing'.[70] The plan disclosed at the end of the summit involves companies committing to pre- and post-deployment testing of advanced AI systems, and recognizes that states have a role to play in that process by investing in public-sector capacity for testing and safety research.

Fourth, the summit's main outcome—the Bletchley Declaration—was signed by 28 states, including the two biggest AI powers, the USA and

[67] Woollacott, E., 'UK announces AI Safety Summit', *Forbes*, 17 Aug. 2023.

[68] British Government Office for Science, *Future Risks of Frontier AI*, Technology & Sciences Insights and Foresight paper, Oct. 2023, p. 2. On the challenges of regulating frontier AI systems see Anderljung, M. and Korinek, A., 'Frontier AI regulation: Safeguards amid rapid progress', Lawfare, 4 Jan. 2024.

[69] AI Safety Summit website.

[70] British Government, Office of the Prime Minister and Department for Science, Innovation and Technology, 'World leaders, top AI companies set out plan for safety testing of frontier AI as first global AI Safety Summit concludes', Press release, 2 Nov. 2023.

China.[71] This can be viewed as a major win for the British government given the level of competition and distrust between the UK and China on this topic.[72] Admittedly, the declaration is relatively weak in terms of political commitments, merely recognizing that frontier AI poses challenges that must be addressed internationally. It does, however, include some normative elements such as that 'AI should be designed, developed, deployed, and used in a manner that is safe, . . . human-centric, trustworthy and responsible'. The declaration also recognizes that 'many risks arising from AI are inherently international in nature' and need to be addressed through international cooperation.[73] In a context marked by increasing competition, such commitments are positive.

Fifth, the summit served as a window of opportunity for some important actors to make key policy announcements. The day before the summit, the US president presented his executive order on 'safe, secure and trustworthy' AI, which establishes commitments and sets new standards across a variety of areas, including testing. [74] That same day the Group of Seven (G7) also presented the conclusion of its 'Hiroshima process', which aims to set guiding principles and a code of conduct for organizations developing advanced AI systems.[75]

The November 2023 summit was the first in a series of meetings to be held every six months. The next two meetings will be held in April and November 2024, in South Korea and France, respectively. The content and contour of these next meetings are not yet known, but the focus will likely remain on risks stemming from frontier AI systems. An open question is whether the future summits will have a more thematic focus given that the Bletchley Declaration flagged three areas of particular concern: (*a*) misinformation, disinformation and hate speech; (*b*) cyber-security; and (*c*) biosecurity.

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

[72] Meacci, L. and Hüsch, P., 'How China and the UK are seeking to shape the global AI discourse', Royal United Services Institute (RUSI) Commentary, 25 Sep. 2023; and Jing Cheng, J. and Zeng, J., 'Shaping AI's future? China in global AI governance', *Journal of Contemporary China*, vol. 32, no. 143 (2023).

[73] Other key outcomes include an AI testing plan, the commission of a report by one of the 'godfathers' of AI, Yoshua Bengio, and the launch of an AI Safety Institute. British Government, 'The Bletchley Declaration' (note 71).

[74] White House, 'President Biden issues executive order on safe, secure and trustworthy artificial intelligence', Briefing Room Fact Sheet, 30 Oct. 2023.

[75] G7, 'Hiroshima Process international code of conduct for advanced AI systems', 30 Oct. 2023; and G7, 'Hiroshima Process international guiding principles for advanced AI system', 30 Oct. 2023.

Conclusions: A pivotal year for the governance of AI at the international level?

It was an eventful year for the governance of AI at the international level. It is probable that 2023 will be remembered as a pivotal year in the political history of AI, in at least three respects.

First, in the CCW, the GGE managed to adopt in its final report language that could form the basis of a two-tiered regulation on AWS. The CCW also adopted a mandate that could mark a potential endpoint for the discussion on AWS in the context of the CCW. Simultaneously, states approved the initiation of a new discussion track under the auspices of the UN General Assembly that could serve as a basis for a future ad hoc process to complement the outcome of the CCW process, or replace it should the latter fail.

Second, states formally acknowledged the need to widen the conversation about AI risks beyond AWS, to cover other ways through which advances in AI may present challenges for international peace and security. In that respect, REAIM and the AI Safety Summit provided different vehicles for discussion of the two main risk pathways. REAIM allows the international community to address the spectrum of humanitarian and strategic risks associated with the potential irresponsible adoption of AI in the military domain. The AI Safety Summit serves as an avenue for discussing the risks stemming from the advances in civilian AI, in particular the misuse of civilian AI in the political and cyber domains, and for other violent purposes.

Third, the conversations concomitantly reached a deeper and higher level. The creation of novel discussion forums allowed states to explore issues at a deeper technical level. At REAIM, for instance, states extensively discussed the problems of transparency, interpretability and bias associated with the use of AI applications based on machine learning. The AI Safety Summit led to extensive discussion and commitment to the testing and evaluation of advanced AI systems. At the same time, these discussions mobilized decision-makers at much higher political levels than ever before. The UN secretary-general and several heads of state engaged personally on the issue. It was also notable that AI was a key point in the bilateral meeting between US President Joe Biden and China's President Xi Jinping in November 2023.[76] These developments provide clear evidence that the issue of AI in 2023 became one of the most important policy issues for the international community.

[76] Hsu, J., 'How the US and China talking AI safety could reduce nuclear war risk', *New Scientist*, 16 Nov. 2023.