EU Non-Proliferation and Disarmament Consortium

Promoting the European network of independent non-proliferation and disarmament think tanks

NON-PROLIFERATION AND DISARMAMENT PAPERS

No. 97 May 2025

LESSONS FROM THE EU ON CONFIDENCE-BUILDING MEASURES AROUND ARTIFICIAL INTELLIGENCE IN THE MILITARY DOMAIN

SOFIA ROMANSKY*

I. INTRODUCTION

Deliberations around the governance of artificial intelligence (AI) in the military domain are rapidly garnering attention in the global arena. Multiple concurrent international and national initiatives have been launched with the goal of establishing core principles and frameworks to guide the development, deployment and use of lethal autonomous weapons (LAWS), as well as AI in the military domain more broadly.[1] There was noteworthy progress in these areas in 2024, such as the first resolution on military AI by the First Committee of the United Nations General Assembly, the second Responsible AI in the Military Domain (REAIM) summit in Seoul, which produced the 'Blueprint for Action' endorsed by over 60 states, and bilateral talks between the United States and China, which led to an agreement on emphasizing human control over nuclear command, control and communications.[2] These developments can be interpreted as reflecting maturing dialogues and converging priorities around core issues, such as accordance with international law, responsibility and accountability, bias and harm mitigation, explainability and traceability, and reliability and governability.[3] While these principles are not universally employed as official terms, they are nonetheless expressed or implicit in the content of ongoing governance processes. There is also prevailing consensus among

[1] While the two focus areas are related, LAWS represent a small but important subset of technologies that are discussed under the umbrella of military applications of AI.

[2] Harjani, M., 'Military AI governance in 2024: One step forward, two steps back', S. Rajaratnam School of International Studies (RSIS), Singapore, 10 Jan. 2025, p. 1; and Harjani, M., 'Parsing the inaugural China-US AI talks', RSIS, 21 May 2024.

[3] Sweijs, T. and Romansky, S., 'International norms development and AI in the military domain', Centre for International Governance Innovation, 4 Sep. 2024, p. 11.

SUMMARY

Integrating artificial intelligence (AI) into the military domain presents a number of significant challenges that have contributed to a deadlock in global governance deliberations. The rapid evolution of AI, its dual-use nature and its impact on the strategic calculations of actors promote the perception that trade-offs are required between security imperatives and ethical and legal considerations. Fortunately, a diverse toolbox of confidence-building measures (CBMs) offers a way forward for governance processes and initiatives by fostering trust and reducing uncertainty. Drawing on lessons from the content of and processes that led up to the European Union AI Act, this report examines how global governance deliberations might benefit from a focus on risks and risk mitigation in order to operationalize high-level principles, as well as multi-stakeholder engagement and investment in an information-based oversight body to ensure that the outcomes of deliberations are relevant and implementable. At the same time, this report also emphasizes the value of EU actorness and the role of European small and middle powers in the global governance arena around AI in the military domain, as well as the need to harmonize civilian and military regulation within the EU. By leveraging CBMs and drawing on the structured regulatory approach of the EU AI Act, global governance efforts can move beyond the current deadlock to foster a more coherent, risk-informed and practical framework for the responsible development, deployment and use of AI in the military domain.

ABOUT THE AUTHOR

Sofia Romansky is a strategic analyst at The Hague Centre for Strategic Studies and the Project Coordinator of the Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM). Her primary research focuses concern the impact of artificial intelligence on the social and military domains, specifically the centrality of the concept of responsibility, issues around narratives and disinformation in online spheres as hybrid threats, and Russia's invasion of Ukraine. Sofia has a Master's Degree in International Relations and Diplomacy from the University of Leiden and the Clingendael Institute.

experts and practitioners around the globe that there is still ample work to be done. A chasm remains between these emerging high-level principles, which are primarily rooted in social norms, or 'intersubjective understandings of appropriateness' that are not necessarily codified but nonetheless recognized by actors, and their practical integration into the AI lifecycle through standards, as well as the application and enforcement of existing international law.[4]

The reluctance to move beyond high-level agreements can be attributed to qualities inherent to AI as a category of technologies, the dynamics of interstate competition and the nature of ongoing governance deliberations.[5]

First, AI in itself is a continually evolving, broad label for a wide range of civilian and military capabilities.[6] In spite of the urgent need for well-evaluated regulation, advances in AI technologies have led experts to argue that 'the current rate of AI development outpaces the rate at which policies can be formulated and adopted', which is also referred to as the 'AI Power Paradox'.[7] Arguably, the core issues that have emerged from the integration of AI into the military domain will not necessarily change as the technology evolves. However, as new capabilities are employed in new contexts, they contribute to further uncertainties about the sufficiency and suitability of existing regulations. At the same time, many AI applications are dual-use, in that technologies can be deployed in civilian and military domains alike, which blurs the boundaries of conventionally disparate regulatory areas.

Second, the models that many AI technologies are based on are often opaque to developers and practitioners alike. Coupled with AI's potentially significant impact throughout the military domain, these qualities motivate increased secrecy around AI-enabled capabilities. This makes it 'harder for policymakers to accurately judge relative military capabilities' or the intentions of rivals.[8]

Finally, the procedural elements of some international deliberations and lack of a clear governance end-game have led to stagnation, where, for example, definitions are discussed to little or no avail.[9] This issue affects both narrow deliberations around LAWS and broader scope deliberations around the military domain alike, detracting from the ability of states to meaningfully engage with each other in global forums. Collectively, these factors contribute to a dilemma experienced by states. Some are unwilling to tie their hands with regulation in contrast to their untethered rivals, which would take full advantage of AI or retain perceived competitive advantages using only lowest common denominator agreements. In the absence of comprehensive, unifying frameworks, such an approach could engender an intensification of competition.[10]

The lens of confidence-building measures

Fortunately, things are not entirely bleak for the governance of AI in the military domain. While the dual-use nature of AI as a category of technologies creates challenges for governance, it also presents opportunities. Notably, dual-use AI is one of the biggest categories procured and deployed by the military, and it therefore faces similar obstacles to integration and regulation as civilian AI. At the same time, even if military AI were procured in-house, the techniques employed would not differ from civilian domain AI. It is therefore possible and even necessary to look to civilian AI regulatory frameworks for ideas on how military AI could be governed.

In a milestone for AI governance, the European Union AI Act was passed in 2024. It constitutes the first legal framework to address risks and regulate the uses of AI and is the leading effort to devise a global

[4] Bode, I., 'Contesting use of force norms through technological practices', *Heidelberg Journal of International Law*, vol. 83, no. 1 (May 2023), p. 42.

[5] The utility of the term 'AI arms race' is widely debated in this context. This paper acknowledges that while the reality of interstate competition around AI in the military domain is not necessarily that of an arms race, perceptions of it as such and consequent attitudes and behaviours still affect the progress of global governance deliberations. Horowitz, M. and Scharre, P., *AI and International Stability: Risks and Confidence-Building Measures* (Centre for a New American Security: Washington, DC, Jan. 2021), p. 4; and Roff, H., 'The frame problem: The AI "arms race" isn't one', *Bulletin of the Atomic Scientists* (blog), 29 Apr. 2019.

[6] Russell, S. J., and Norvig, P., *Artificial Intelligence: A Modern Approach* (Pearson, 2016).

[7] Sweijs and Romansky (note 3), p. 5.

[8] Horowitz and Scharre (note 5), p. 6; and Michel, A. H., *The Black Box, Unlocked: Predictability and Understandability in Military AI* (UNIDIR: Geneva, 2020).

[9] Schmitt, L., 'Mapping global AI governance: A nascent regime in a fragmented landscape', *AI and Ethics*, vol. 2, no. 2 (May 2022), p. 9.

[10] Soare, S. R., 'European military AI: Why regional approaches are lagging behind', eds M. Raska and R. A. Bitzinger, *The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities, and Trajectories* (Routledge: New York, 2023), p. 78; and Maas, M. M. and Villalobos, J. J., 'International AI institutions: A literature review of models, examples, and proposals', AI Foundations report 1, *SSRN Scholarly Paper*, Rochester, NY, 22 Sep. 2023, p. 24.

approach to AI technology regulation standards.[11] In addition to its direct implications for dual-use AI technologies, the AI Act also offers valuable content and process-based lessons for the governance of AI in the military domain, such as through confidence-building measures (CBMs).[12] This report holds that CBMs provide an informative lens through which AI regulation should be considered, to soberly observe the dynamics at play in the current stage of global governance deliberations and suggest a way forward. The EU AI Act serves as a key case study.

CBMs are a toolbox of strategies that states and non-governmental stakeholders have at their disposal to formulate shared governance frameworks while addressing concerns related to competitive pressures. CBMs encompass an array of methods that increase mutual understanding and trust, clarify intentions and reduce misperceptions, miscalculations and the risk conflict.[13] These tools are especially valuable in less established deliberations around contentious subjects that are primarily geared to finding common ground between different perspectives, as is the case in current processes around AI in the military domain.[14]

This paper focuses on two key lessons for the development of international CBMs that can be drawn from the EU AI Act. First, the EU AI Act can help to inform the broad priorities of international governance deliberations. Specifically, the AI Act emphasizes the importance of: (*a*) operationalizing high-level principles; (*b*) directing the focus to risks and risk-mitigation; (*c*) facilitating multi-stakeholder engagement; and (*d*) investing in the creation of an oversight body. Second, lessons from the processes behind the AI Act can inform how the EU and its member states should interact with global governance deliberations as important international actors. It demonstrates the need to embrace the relevance of the EU and European small and middle powers in global governance deliberations and to work to increase cohesion between EU civilian and military policies. The recommendations reflect a widespread interpretation that the EU AI Act represents an avenue through which the EU can exercise its normative and regulatory power. In practical terms, the means through which the AI Act was negotiated and the ways in which values have been embedded in it can be used as inspiration for CBMs to be employed around military AI at the international level.

Section II provides a brief introduction to the concept of CBMs and their importance for the global governance of AI in the military domain. Section III reviews the role that CBMs have played in ongoing international deliberations. Section IV analyses the salience of CBMs in the processes and content of the EU AI Act. Section V concludes by identifying key lessons that can be drawn from the Act and makes recommendations for the EU and its members states, and on global governance deliberations more broadly.

II. CONCEPTUALIZING CONFIDENCE-BUILDING MEASURES AND AI IN THE MILITARY DOMAIN

Confidence-building measures, or tools that aim to help manage security dilemmas and prevent escalation while fostering trust between states and other actors, are neither new nor unique to conversations around AI. The concept of CBMs emerged during the cold war as a collection of mechanisms aimed at fostering the 'exchange of information, notification, and observation, on a voluntary basis, of major military activities'.[15] In this context, CBMs were initially employed as part of arms control around weapons of mass destruction. They played a vital role in preventing strategic miscalculations and laid the foundations for treaties, particularly between the nuclear powers.[16] Over time, the issue areas that rely on CBMs have expanded to include 'emerging disruptive technologies' (EDTs).

While there is no universally accepted definition of CBMs, most practices revolve around the various ways in which actors can voluntarily contribute to crisis aversion and the maintenance of good relations.[17] To add nuance to conversations around CBMs for AI specifically, one typology distinguishes between capability-based and behaviour-based

[11] Csernatoni, R., 'Governing military AI amid a geopolitical minefield', Carnegie Endowment for International Peace, 17 July 2024.

[12] Csernatoni, R. , 'Weaponizing innovation: Mapping Artificial Intelligence-enabled security and defence in the EU', EUNPDC, 6 July 2023, p. 13; and Ronnback, R., 'Challenges of governing AI for military purposes and spill-over effects of the AI Act', Futurium, European AI Alliance, 27 Feb. 2023.

[13] United Nations Office for Disarmament Affairs (UNODA), *Securing Our Common Future: An Agenda for Disarmament* (UNODA: New York, 2018), p. 11; and Puscas, I., *Confidence-Building Measures for Artificial Intelligence: A Multilateral Perspective* (UNIDIR, 2024), p. 10.

[14] Horowitz, M. C., Kahn, L. and Mahoney, C., 'The future of military applications of artificial intelligence: A role for confidence-building measures?', *Orbis*, vol. 64, no. 4 (Jan. 2020), pp. 528–43.

[15] Puscas (note 13), p. 8; and Desjardins, M-F., 'In search of a theory: Developing the concept', *Adelphi Papers*, vol. 36, no. 307 (Dec. 1996), p 7.

[16] Horowitz and Scharre (note 5), p. 5.

[17] United Nations Office for Disarmament Affairs (note 13), p. 11.

measures. Capability-based CBMs focus on physically tangible, technical and structural security-enhancing mechanisms, such as arms control with the aim of curbing proliferation, data exchange and verification agreements. Behaviour-based CBMs underpin diplomatic and institutional practices that can enhance communication and assuage doubts between actors through, for example, crisis hotlines and structured dialogues.[18] Both types of measure are valuable not only for the outcomes they can achieve, but also as processes in and of themselves. Bringing actors together to deliberate on CBMs in open-ended processes often constitutes a first step towards opening up and maintaining diplomatic channels of communication on critical issues.[19] Consequently, it has been acknowledged that CBMs are especially useful in contexts where contentious disputes over war-related technologies might create uncertainty, soft law frameworks provide direction without legal enforcement, and social norms inspire trust and cooperation among actors for a collective good.[20]

Artificial intelligence and uncertainty

CBMs can play a significant role in global governance deliberations around AI in the military domain. Specifically, they can help to mitigate various sources of uncertainty intrinsic to military AI applications, such as their ability to affect regional balances of power, their dual-use nature and the involvement of various actors throughout the AI lifecycle, as well as and their technical opacity.

First, it is necessary to acknowledge that many military powers are making significant investments in AI because of the projected efficiency gains offered by AI applications, especially in situations where human resources are limited. At the same time, however, these potential gains contribute to the 'considerable uncertainty about the extent to which military AI will alter global and regional power balances'. The proliferation of AI technologies could act as a capability equalizer, especially in instances where states lacked manpower.[21] In addition, as the category of technologies continues to rapidly develop and evolve, it could reshape the character of conflict and the capabilities required by modern militaries, while also increasing the speed and scale of military activities by processing greater quantities of diverse data.

Second, AI constitutes a category of technologies that blurs the boundaries of the military domain, due to the dual-use nature of many of its applications and the variety of actors involved throughout the AI lifecycle. AI innovation is currently being spearheaded by private sector actors, which obscures evaluations of the level of innovation and technological maturity among states. The dual-use nature of AI technologies contributes an additional layer of uncertainty as it becomes difficult to determine whether a technology or system was created for military or civilian purposes.[22] While some AI technologies will certainly be developed in more bespoke ways, this does not preclude the possibility that militaries might procure capabilities that fundamentally were not developed for military purposes, specifically in areas such as border control where military and civilian domains overlap.[23]

Finally, AI systems are often 'black boxes' for practitioners and developers alike, where it is unclear how an output was achieved. There is consequently a degree of output-based uncertainty that arises from adaptable, self-learning models.[24] Thus, even if these technologies were developed in transparent ways, with clarity about their intended purpose, unexpected outcomes could arise due to the specific AI techniques employed. In turn, these systems can contribute to the risk of inadvertent conflict escalation due to the emergence of unpredictable or erroneous outputs that could be misconstrued by adversaries or lead to miscalculations.[25]

Ultimately, the convergence of these factors in a new software-driven warfare paradigm complicates the assessment of an adversary's capabilities and intentions. Combined with the confidential nature of

[18] Puscas (note 13), p. 10; and Horowitz, M. C., 'Artificial intelligence, international competition, and the balance of power', *Texas National Security Review*, vol. 1, no. 3 (May 2018).

[19] Imbrie A. and Kania, E., 'AI safety, security, and stability among great powers: Options, challenges, and lessons learned for pragmatic engagement', *Center for Security and Emerging Technology*, Dec. 2019, p. 12.

[20] Horowitz, Kahn and Mahoney (note 14).

[21] Sweijs and Romansky (note 3), p. 7.

[22] Carrozza, I., Marsh N. and Reichberg, G., *Dual-Use AI Technology in China, the US and the EU: Strategic Implications for the Balance of Power* (Peace Research Institute Oslo: Oslo, 2022), p. 9.

[23] Carrozza, Marsh and Reichberg (note 22).

[24] Verbruggen, M. and Boulanin, V., *Mapping the Development of Autonomy in Weapon Systems* (Stockholm: Stockholm International Peace Research Institute, 2017).

[25] Csernatoni and EUNPDC (note 12), p. 3; and Shoker, S. et al., 'Confidence-building measures for Artificial Intelligence', Workshop proceedings, 3 Aug. 2023.

military information that has direct implications for national security, this is likely to make military actors reluctant to engage in information-sharing accords.[26]

It is of paramount importance that militaries only employ technologies that are maximally reliable, predictable and governable, to ensure that they perform as anticipated. In addition, how systems function needs to be sufficiently understood by human operators as a prerequisite for establishing accountability under international law.[27] However, the dynamics of uncertainty around the role and reach of AI technologies could drive militaries to make decisions that prioritize keeping up with adversaries rather than mitigating the above-mentioned risks.[28]

Artificial intelligence and confidence-building measures

Although these dynamics are troubling, they can be managed and mitigated through CBMs, which can communicate expectations around these technologies under international law and aid de-escalation in instances of AI-based miscalculations.[29] To achieve this, CBMs for AI in the military domain would need to account for the implications of an ongoing transition from hardware-driven capabilities to those that integrate algorithms.

First and foremost, it is important to note that this shift has significant implications for the use of verification and capability-based CBMs. In contrast to conventional arms control agreements, which rely on physical inspections and satellite imagery, for example, the presence, maturity and intended purpose of AI systems are more difficult to assess, even when integrated into physical systems.[30] While a discussion on verification is beyond the scope of this paper, certain lessons can be drawn from the governance of the cyber domain, specifically in the light of the digital nature of many AI interfaces, platforms and outputs. For example, existing protocols for verifying data safety practices necessarily apply to foundational AI models, and demonstrate that it is possible to establish guidelines for rapidly proliferating, dual-use, decentralized technologies.[31] However, the additional autonomous and emergent qualities of AI applications mean that cybersecurity and data regulation cannot offer an exact regulatory template for AI.[32] Concurrently, verification is not a strictly necessary component of a robust CBM framework. CBM arrangements based on information exchange can achieve desired outcomes, even in the absence of verification, by focusing on behaviour-based CBMs that aim to maximize trust and transparent communication. In turn, these new CBMs would need to account for how a primarily software-based collection of technologies creates new types of uncertainty to be addressed.

Second, unlike other technologies, such as those that underly weapons of mass destruction, the development of AI systems is primarily being led by the private sector, which introduces additional convoluted chains of responsibility. Thus, an approach is needed that can account for the multi-faceted functional, legal, ethical and operational risks that stem from the private sector development of these AI systems and their use in the military domain.[33] Such a governance framework should include risk mitigation measures that ensure that AI deployment in military contexts remains aligned with innovation goals and governance standards, while still allowing states to access military AI capabilities without compromising security.[34] By drawing lessons from a structured regulatory approach such as the EU AI Act, CBMs for AI in the military domain can be refined to balance trust-building with national security interests, thereby ensuring that governance frameworks are constructed to be agile in an evolving technological landscape.

[26] Trabucco, L. and Maas, M. M., *Technology Ties: The Rise and Roles of Military AI Strategic Partnerships* (SSRN Scholarly Paper: Rochester, NY, Nov. 2023).

[27] Kraska, J., 'Command accountability for AI weapon systems in the law of armed conflict', *International Law Studies*, vol. 97, no. 1 (Jan. 2021), pp. 407–47; and Novelli, C., Taddeo, M. and Floridi, L., 'Accountability in artificial intelligence: What it is and how it works', *AI & SOCIETY*, 7 Feb. 2023.

[28] Horowitz and Scharre (note 5), p. 5; and Kwik, J. and Van Engers, T., 'Algorithmic fog of war: When lack of transparency violates the law of armed conflict', *Journal of Future Robot Life*, vol. 2, no. 1–2 (Jan. 2021), pp. 43–66.

[29] Horowitz and Scharre (note 5), p. 7; and Konaev, M. and Lohn, A., 'Confidence-building measures for artificial intelligence', Workshop proceedings, Center for Security and Emerging Technology, 4 Aug. 2023, p. 1.

[30] Scharre, P. and Lamberth, M., 'Artificial intelligence and arms control', 22 Oct. 2022.

[31] Jurić, M., *Legal Aspects of Military and Defence Applications of Artificial Intelligence Within the European Union* (Central European Academic Publishing: Budapest, 2024), p. 416.

[32] Ronnback (note 12).

[33] Cooper, S., Copeland, D. and Sanders, L., 'Methods to mitigate risks associated with the use of AI in the military domain', ed. Schraagen, J. M., *Responsible Use of AI in Military Systems* (Chapman and Hall/CRC: New York, 2024), p. 128.

[34] Horowitz (note 18).

Table 1. Overview of existing international governance initiatives around LAWS and AI in the military domain

Initiative	Years active	Key outputs	Main focus	Representation and actors involved
United Nations General Assembly First Committee Resolution on Artificial intelligence in the military domain and its implications for international peace and security[a]	2024–present	Resolution (2024)	**AI in the military domain**, state responsibility	**Global** 165 in favour to 2 against (Democratic People's Republic of Korea, Russian Federation) with 6 abstentions (Belarus, Ethiopia, Iran, Nicaragua, Saudi Arabia, Syria)
Economic Community of West African States Conference on Autonomous Weapons Systems	2024	Freetown Communique (2024)[b]	**Autonomous weapon systems**, legally binding instruments	**Regional** 15 ECOWAS member states
Responsible AI in the Military Domain Summit Series and Regional Consultations	2023–present	REAIM Call to Action (2023)[c] REAIM Blueprint for Action (2024)[d]	**AI in the military domain**, responsible AI, multi-stakeholder engagement	**Global** Led by the Republic of Korea, the Netherlands, Singapore, Kenya and the United Kingdom 57 signatories as of 2023 64 signatories as of 2023
Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy[e]	2023–present	Political Declaration (2023)	**Artificial intelligence and autonomy in the military domain**, exchange of practices and capacity building	**'Western allies' of the United States** Led by the United States 58 signatories as of 2023
Caribbean Community Declaration on Autonomous Weapons Systems[f]	2023	Declaration (2023)	**Autonomous weapons systems**, legally binding instruments	**Regional** 15 CARICOM member states
Belen Communiqué of the Latin American and the Caribbean Conference of Social and Humanitarian Impact of Autonomous Weapons[g]	2023	Communique (2023)	**Autonomous weapons systems**, legally binding instruments	**Regional** 33 states from Latin American and the Caribbean, led by Costa Rica
NATO Artificial Intelligence Strategy	2021	Revised Strategy (2024)[h] Strategy (2021)[i]	**AI in the military domain**, responsible AI	**Alliance-based** 32 NATO member states and 40 non-member states and international organizations
Group of Governmental Experts on Lethal Autonomous Weapons	2016–present	Rolling Text (2024)[j] Guiding Principles (2019)[k]	Lethal autonomous weapons systems	**Global** High Contracting Parties and non-states parties to the CCW, international organizations and non-governmental organizations

[a] United Nations, General Assembly, First Committee Resolution on artificial intelligence in the military domain and its implications for international peace and security, 19 Oct. 2024.

[b] Economic Community of West African States, Conference on Autonomous Weapons Systems, Freetown Communiqué, 18 Apr. 2024.

[c] Responsible AI in the Military Domain (REAIM), Call to Action, 16 Feb. 2023.

[d] Responsible AI in the Military Domain (REAIM), Blueprint for Action, 11 Sep. 2024.

[e] US Department of State, Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.

[f] Caribbean Community Declaration on Autonomous Weapons Systems, 6 Sep. 2023.

[g] Latin American and the Caribbean Conference of Social and Humanitarian Impact of Autonomous Weapons, Belen Communiqué, 24 Feb. 2023.

[h] NATO, 'Summary of NATO's revised Artificial Intelligence (AI) strategy', 10 July 2024.

[i] NATO, 'Summary of the NATO Artificial Intelligence (AI) strategy', 22 Oct 2021.

[j] Group of Governmental Experts on Lethal Autonomous Weapons, Rolling text, 8 Nov. 2024.

[k] Group of Governmental Experts on Lethal Autonomous Weapons, Annex III, Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, CCW/MSP/2019/9.

III. CONFIDENCE-BUILDING MEASURES IN DELIBERATIONS ON GLOBAL GOVERNANCE OF AI IN THE MILITARY DOMAIN

To help determine the next steps for CBMs as part of deliberations on the governance of AI in the military domain, it is useful to take stock of what is already in place. As of 2024, at least eight initiatives were addressing the issue in conferences, declarations, resolutions and strategies (see table 1).

The longest running ongoing initiative is the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE on LAWS), which operates under the 1983 Convention on Certain Conventional Weapons (CCW). The CCW has been active around emerging technologies in the area of LAWS since 2014. It convened an informal group of experts before the GGE was established in 2016.[35] A 2019 GGE report on LAWS outlined 11 guiding principles that have become a benchmark for further governance efforts.[36] While the GGE on LAWS has a narrower scope than some other governance initiatives, LAWS constitutes a core issue area in whole-of-military-domain discussions.[37] In recent years, discussions in the GGE have looked to expand their focus, highlighting the inextricable link between LAWS and other AI systems in the military domain.[38]

NATO's 2021 Artificial Intelligence Strategy, which was revised in 2024, was another influential development. It specifies six principles for the responsible use of AI in defence to guide its work and that of its member states.[39] In 2023, the Caribbean Community (CARICOM) published a Declaration on Autonomous Weapons Systems, which was followed by the Belen Communiqué from the 'Latin American and the Caribbean Conference of Social and Humanitarian Impact of Autonomous Weapons'. These are two key regional initiatives that stress the importance of the creation of legally binding instruments to regulate the use of LAWS. The 2023 REAIM Summit and the 2023 Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy are among the most wide-reaching initiatives to date. As part of the REAIM initiative, regional consultations were held throughout 2024, culminating in the second REAIM Summit in Seoul. The 2024 Freetown Communique by the Economic Community of West African States echoes the urgent calls for legally binding measures on autonomous weapon systems made in the CARICOM Declaration and the Belen Communiqué. The final key development in 2024 was the first resolution on military AI passed by the First Committee of the United Nations General Assembly, which achieved near-universal endorsement.

The increasing number of governance efforts has arguably created risks of fragmentation, duplication of effort and forum shopping, spreading stakeholders too thin and leading to potential inconsistencies in governance approaches.[40] Fragmentation could arise from the diverse mandates of the above initiatives. The processes that contributed to the UN resolution, the GGE on LAWS and the NATO strategy are based on the UN Charter, the Convention on Certain Conventional Weapons and the North Atlantic Treaty, respectively, which means that they are embedded into existing structures. The REAIM and the Political Declaration, by contrast, are newer initiatives that aim to contribute to high-level discussions, and thus have no established processes. Finally, the CARICOM Declaration and the Belen and Freetown communiqués reflect one-off regional calls for action and legally binding instruments that would stem from different sources of hard law.

While all the initiatives are worth noting, certain avenues and collections of actors will undoubtedly have a greater impact on global governance than others. A core group of primarily European states participates in almost all the international initiatives, while Latin American and Caribbean states have demonstrated a preference for maintaining a regional focus in their engagement or remaining on the periphery of larger initiatives. At the same time, key states in North America, Europe and East Asia remain at the forefront of innovation, which contributes to a capability gap with the global majority. Thus, in deliberations, some

[35] Geneva Internet Platform, 'GGE on Lethal Autonomous Weapons Systems', Digital Watch Observatory, Digital Watch (blog), 6 Feb. 2025.

[36] United Nations, 'Meeting of the High Contracting Parties to 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', Geneva, 13–15 Nov. 2019.

[37] Blanchard, A. et al., 'Dilemmas in the policy debate on Autonomous Weapon Systems', SIPRI topical backgrounder, 6 Feb. 2025.

[38] Blanchard et al. (note 37).

[39] NATO, 'Summary of NATO's revised Artificial Intelligence (AI) strategy', 10 July 2024.

[40] Murphy, H. and Kellow, A., 'Forum shopping in global governance: Understanding states, business and NGOs in multiple arenas', *Global Policy*, vol. 4, no. 2 (2013), pp 139–49; and Pankakoski, T. and Vihma, A., 'Fragmentation in international law and global governance: A conceptual inquiry', *Contributions to the History of Concepts*, vol. 12, no. 1 (June 2017), pp. 22–48.

states are mostly concerned with the governance of existing capabilities while others prioritize capability development or the mitigation of risks that stem from unequal access to technologies.[41] This tension within global governance is further evident in the emergence of smaller alliance-based collaborations around AI. This is seen in initiatives such as the 2022 Security Pact between Australia, the United Kingdom and the United States (AUKUS), in which Pillar II focuses on emerging technologies such as AI, and in partnerships between China and Russia.[42] In this way, global views may become fractured as states find it more advantageous to engage with like-minded allies than to attempt to bridge ideological divides. At the same time, several African and Asian states have not been involved in any of the above-mentioned processes, raising concerns about the representativeness of purportedly global initiatives.[43] Finally, while many states attend consultations or summits, such as those of REAIM, their involvement does not always translate into endorsement of the outcome documents.

A sense of urgency and a shared vocabulary

Nonetheless, the recent boom in global governance initiatives can also be interpreted as underlining an international sense of urgency, a growing recognition of the security implications of military AI and the need for international coordination through a shared vocabulary.

Despite the different focus areas between LAWS or AI in the military domain more broadly, and the mandates of concurrent governance initiatives, their mere existence can be considered a kind of behaviour-based CBM. All eight initiatives involve or have involved international meetings that serve as platforms for dialogue, which signals to the global community that the integration of AI into the military domain is an issue that transcends state boundaries. Furthermore, common traits among these initiatives indicate the emergence of informal behaviour-based standards for CBMs in military AI governance. Not only REAIM, but also the consultations on the Political Declaration and the work of the GGE on LAWS incorporate a degree of multi-stakeholder engagement between state delegations, experts and the private sector in their processes. This aligns with the arguments made in section II regarding why CBMs are particularly valuable tools for AI in the military domain. Engagement between diverse stakeholders is needed where a variety of actors, beyond state-linked institutions, contribute to technological innovation. While there are clear trends towards multi-stakeholderism, deep cooperation between rival states remains limited due to national security concerns and the risk of technology proliferation. This reluctance in turn affects the ability and willingness of technical experts to engage in meaningful collaborations.[44]

Global governance deliberations have also positively contributed to a shared vocabulary for discussing risks, opportunities and social norms related to military AI. The proliferation of terms such as 'responsible AI' and 'meaningful human control', while still definitionally fuzzy, has helped to encapsulate the core issues in this area.[45] Despite the differences in fundamental, semantic approaches, a shared understanding is important not only because of the variety of actors involved throughout the AI lifecycle, but also to help counter misconceptions about AI governance and contribute to a sense that conflicting interests are not insurmountable obstacles.[46] For instance, there is a frequent misconception that mechanisms designed to limit harm could hinder the peaceful use of or innovation in AI technologies. However, governance, including CBMs, and technological development are not inherently mutually exclusive.[47] While several states have noted that CBMs should function in tandem with legally binding agreements, rather than act as a replacement, there appears to be a distinct value in CBMs at the current stage of deliberations and beyond. In fact, CBMs continue to be useful both as mechanisms in their own right and when accompanied by legally binding measures that may have emerged as a result of engagement around CBMs.[48]

[41] Canfil, J. K. and Elsa, K. B., 'Mapping state participation in military AI governance discussions', eds J. B. Bullock et al., *Oxford Handbook of AI Governance* (Oxford University Press: Oxford, 2024).

[42] Trabucco and Maas (note 26), p. 3; and Sweijs and Romansky (note 3), p. 21.

[43] See Figure 1. For a breakdown of the membership of the international governance initiatives focused on LAWS or AI in the military domain, see Sweijs and Romansky (note 3), p. 10.

[44] Sweijs and Romansky (note 3), p. 20; and Maas and Villalobos (note 10), p. 24.

[45] Tigard, D. W., 'Responsible AI and moral responsibility: A common appreciation', *AI and Ethics*, vol. 1, no. 2 (May 2021), pp. 113–17.

[46] Hass, R. and Kahl, C., 'Laying the groundwork for US-China AI dialogue', Brookings, 5 Apr. 2024.

[47] Sweijs and Romansky (note 3), p. 20; and Greene, N., 'The EU AI Act could hurt military innovation in Europe', Encompass, Jan. 2024.

[48] Puscas (note 13), p. 6.

Nonetheless, only a minority of governance initiatives explicitly incorporates CBMs into their official documentation and outcomes, by stipulating mechanisms within their governance frameworks that would fall under the CBM conceptual umbrella. Among the multilateral initiatives mentioned, three have explicitly emphasized exchange of practices as a normative focus area and behaviour-based CBMs: the REAIM Summit Call to Action, the Communiqué of the Latin American and the Caribbean Conference of Social and Humanitarian Impact of Autonomous Weapons and the CARICOM Declaration on Autonomous Weapons. The emphasis on specifically behaviour-based CBMs in the CARICOM Declaration and Latin American and the Caribbean Communique bolsters the perception that these are a means through which regions can strengthen their position in deliberations and regulatory processes. Meanwhile, the REAIM Call to Action mentions the need to increase the 'exchange of lessons learnt regarding risk mitigation practices and procedures', to exchange information between all stakeholders on responsible AI and to exchange good practices 'to increase the mutual comprehension of states' national frameworks and policies', as well as the need 'for all states, especially developing countries, to benefit from the opportunities and to address the challenges and risks'.[49] The centrality of information exchange is no surprise as it reflects the nascent stage of discussions around AI in the military domain and the tangible pressure felt by actors to reduce uncertainty.

In the deliberations around the governance of AI in the military domain, CBMs are gaining prominence as tools for risk mitigation, norm diffusion and trust-building through ongoing global governance initiatives.[50] Despite these efforts, however, several critical gaps remain in the operationalization of CBMs within global governance deliberations. For instance, there is no coherent agreement on the issues around capability-based CBMs and the risks that arise from the dual-use nature of AI as a category of technologies.[51] The question therefore arises whether existing CBM initiatives are the best fit for their intended purpose or further refinement of strategies is necessary. As a regional initiative, the EU AI Act helps to demonstrate how structured regulatory frameworks can contribute to both capability- and behaviour-based CBMs in the governance of military AI. In addition, while primarily focused on civilian applications, the EU AI Act can provide specific lessons on how to operationalize risk-based governance, principles of ethical AI and the identification and allocation of responsibilities among various actors. By facilitating these elements, CBMs as part of global governance efforts can serve as a bridge between voluntary cooperation and potentially binding mechanisms for AI in the military domain.

IV. THE EU AI ACT AND CONFIDENCE-BUILDING MEASURES

The 2024 EU AI Act is a major development in AI governance. The Act constitutes the first legislative framework for AI in the civilian domain and sets a precedent for the regulation of civilian AI in a structured and binding way. While experts are quick to note that the act explicitly does not legislate on the use of AI in the military domain, as defence policy decisions remain with the EU member states, its content has direct implications for dual-use AI applications.[52] The processes through which the act was developed also provide valuable lessons for global governance.[53] Ultimately, this interpretation of the role of the EU AI Act is aligned with the EU's broader ambition to extend the 'Brussels effect' to AI, influencing global governance by encouraging external actors to align with EU regulatory norms, but also demonstrating the relevance and utility of the EU institutions or such norms in this sphere.[54]

Deliberations around the EU AI Act began in 2018, following publication of the EU Coordinated Plan on AI. This outlined the EU's research and development priorities and laid the foundations for an ethical framework around AI that would ensure that its uses were in accordance with European values and human rights.[55] Subsequently, the EU established the High-Level Expert Group on AI (HLEG), which was tasked with elaborating core principles to guide AI

[49] Government of the Netherlands, 'REAIM 2023 Call to Action', 16 Feb. 2023.

[50] Puscas (note 13).

[51] Saunders, L. and Copeland, C., 'Developing an approach to the legal review of Autonomous Weapon Systems', *ILA Reporter*, International Law Association, Australia, 27 Nov. 2020.

[52] Soare (note 10); Csernatoni (note 11); and Fanni, M., 'Why the EU must now tackle the risks posed by military AI', CEPS, 8 June 2023.

[53] Csernatoni (note 11); and Greene (note 47).

[54] Creutz, K. et al., *The EU and Military AI Governance: Forging Value-based Coalitions in an Age of Strategic Competition*, Finnish Institute of International Affairs (FIIA) Working Paper (FIIA: Helsinki, 2024), p. 16.

[55] European Commission, 'Coordinated Plan on Artificial Intelligence: Shaping Europe's Digital Future', 7 Feb. 2025.

Table 2. Summary of the EU AI Act's tiered, risk-based approach

Tier	Obligations	Use cases
Unacceptable risk	Prohibited	Social scoring, mass surveillance, manipulation of behaviour to cause harm
High risk	Conformity assessment • Risk-mitigation • High-quality data sets • Clear user information • Human oversight	Access to employment, education and public services, safety components of vehicles, law enforcement
Limited risk	Transparency • Labelling of AI	Impersonation, chatbots, emotion recognition, biometric categorization, deep fakes
Minimal risk	Voluntary codes of conduct	Uses not covered by other categories, such as spam filters, AI-enabled video games etc.

Source: European Commission, 'AI Act enters into force', News article, 1 Aug. 2024.

legislation and related regulatory measures.[56] These efforts produced two keystone documents.[57] These endeavours were complemented by the work of the Global Tech Panel (GTP), which bridged conversations between the civilian and security domains in relation to the EU Common Security and Defence Policy (CSDP).[58] Consolidating this initial work, the European Commission published the White Paper on Artificial Intelligence in 2020, which proposed a risk-based approach to AI regulation and served as a launchpad for the negotiations on a comprehensive EU AI Act that commenced in 2021.[59] The Act underwent extensive negotiation and a final agreement was reached at the end of 2023. Under its provisions, the AI Act will be fully implemented by 2026. Key steps will require the mandatory compliance of private sector actors and the work of the European AI Board (EAIB) as an oversight body.

The EU AI Act provides for a tiered, risk-based approach that divides uses of AI into four classifications: unacceptable, high risk, limited risk and minimal risk (see table 2). Uses of AI that pose unacceptable risks, such as those that manipulate human behaviour or directly impinge on human rights, are prohibited. The other three tiers are subject to different levels of regulatory oversight.[60] As part of risk mitigation, the Act operationalizes several core principles that have become commonplace under the umbrella of 'responsible AI', such as trustworthiness and transparency, accountability and human oversight.[61] Transparency requirements make systems subject to explainability assessments to ensure that human operators are able to understand how and why AI systems produce certain outputs. These standards are most stringent for high-risk systems. At the same time, obligations related to accountability are placed on developers, which are expected to be able to provide authorities with sufficient documentation on research and testing stages on request. Finally, the principle of human oversight acts as a red thread throughout the Act, emphasizing that humans must remain in charge of critical decisions.[62]

The Act acknowledges that uses of AI in military contexts remain under the jurisdiction of individual EU member states and are subject to public international law, which constitutes the legal framework best suited to the regulation of AI technologies 'in the context of the use of lethal force'.[63] Thus, AI applications developed exclusively for military purposes are not covered by the legislation. However, AI systems developed for military purposes but later repurposed for civilian or dual-use applications fall under the

[56] Dai, S. and Song, L., 'Balancing security and regulation: The EU's conundrum in military AI governance', *Security Science Journal*, vol. 5, no. 3 (2024), p. 19.

[57] European Commission, 'Ethics guidelines for trustworthy AI: Shaping Europe's digital future', 8 Apr. 2019; and Dervishaj, J., European AI Alliance, 'Policy and investment recommendations for trustworthy artificial intelligence', 26 June 2019.

[58] Dai and Song (note 56), p. 20.

[59] European Commission, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, COM(2020) 65 final, 19 Feb. 2020.

[60] Ronnback (note 12).

[61] Siegmann, C. and Anderljung, M., 'The Brussels effect and Artificial Intelligence: How EU regulation will impact the global AI market', arXiv, 23 Aug. 2022.

[62] Baker, T., 'The EU AI Act: A primer', Center for Security and Emerging Technology (blog), 26 Sep. 2023; and Zhong, H., 'Implementation of the EU AI Act calls for interdisciplinary governance', *AI Magazine*, vol 45, no. 3 (2024), pp. 333–37.

[63] Jurić (note 31), p. 404.

scope of the Act.[64] In this way, the AI Act articulates how AI inherently blurs the boundaries between the military and civilian domains, which has been a core challenge for global governance.[65] At the same time, the Act functions through an understanding that 'these technologies have implications that transcend national borders, making coordinated governance and oversight necessary [and that] existing frameworks of defense cooperation at the EU level, such as the European Defence Fund, already demonstrate EU member states' capacity for coordinated and integrated efforts to address complex security and defense challenges'.[66]

Lack of a unified EU AI strategy

Nonetheless, the EU AI Act and the EU's approaches to AI more broadly have not gone without criticism. In the international realm, the EU faces a fundamental tension between its self-designated identity as a normative power and the inescapable security implications of AI regulation, precisely because of the dual-use nature of the technology.[67] This issue is exacerbated by the lack of coherence between civilian and security approaches to AI in the EU, which leaves it without a clear framework that 'links technological power to strategic autonomy in terms of operational advantage against and competitiveness with other rival great powers... [their] innovation efforts and other international actors' geopolitical needs'.[68] In both the civilian realm and the security realm there are diverging priorities between the European Parliament, the European Commission, the various AI-based initiatives of Permanent Structured Cooperation (PESCO), the European Defence Fund (EDF) and the CSDP, as well as the involvement of EU member states in NATO.[69] There is also an 'inconsistency between the European Commission's position on excluding military AI from its emerging AI policy' and 'EU policy initiatives targeted at supporting military and defence elements of AI' at the EU level.[70] Consequently, while the AI Act has been referred to as giving the EU a regulatory first-mover advantage, other policies within the EU could undermine the potency of such influence.

For example, the European Parliament has called for a total ban on LAWS as part of its commitment to shaping military AI governance through principles of restraint and accountability, which aligns with the core principles of the AI Act. However, this reflects a disconnect between normative idealism and what is practically feasible or desirable at the member state policy level. While states such as Austria have been vocal about supporting a total ban on LAWS, states such as France advocate the importance of AI for strategic autonomy.[71] This is not necessarily surprising, as larger, defence-industrial nations push for greater AI investment and innovation, while smaller states are still reluctant. Although the EU has been able to reach a level of agreement on issues such as procurement and security of supply, member states continue to invoke article 346 of the Treaty on the Functioning of the European Union, which allows for exemptions from directives where matters of national security are concerned. These tensions raise questions about the coherence of EU practices and highlight the challenge of balancing regulatory ambitions, strategic autonomy and security imperatives in its civilian and military AI governance frameworks.

The utility of CBMs in AI governance

Despite certain criticisms of the content of and process behind the EU AI Act, various lessons can be learned about the utility of CBMs in AI governance. The process of negotiating and adopting the EU AI Act itself functioned as a behaviour-based CBM by fostering trust, transparency and inclusivity among stakeholders. Starting from the high-level ethical principles in the 2018 EU Coordinated Plan on AI, a set of common priorities for AI governance was established to guide future regulation.[72] The creation of the HLEG and the GTP facilitated multi-stakeholder engagement by including technical experts, industry leaders, policymakers and security professionals in the regulatory discussions.[73] The Act's extensive,

[64] Jurić (note 31), p. 406.

[65] Lingevicius, J., 'Military artificial intelligence as power: Consideration for European Union actorness', *Ethics and Information Technology*, vol 25, no. 1 (Feb. 2023), p 5.

[66] Csernatoni (note 11).

[67] Lingevicius (note 65), p. 5.

[68] Soare (note 10), p. 78.

[69] Lingevicius (note 65), p. 1.

[70] Lingevicius (note 65), p. 18.

[71] Badell, D. and Schmitt, L., 'Contested views? Tracing European positions on Lethal Autonomous Weapon Systems', *European Security*, vol. 31, no. 2 (Apr. 2022), pp. 242–61; Dai and Song (note 56), p. 19; and Johansson-Nogués, E., Vlaskamp, M. and Barbé, E. (eds), *European Union Contested: Foreign Policy in a New Global Context*, Norm Research in International Relations (Springer: Cham, IL, 2020).

[72] Lingevicius (note 65), p. 13.

[73] Dai and Song (note 56), pp, 19–20.

multi-year negotiation process (2021–2023) provided a mechanism for consensus-building among EU institutions and member states, which facilitated cooperation in the context of a high-stakes governance challenge.[74] This required continuous clarification of how the AI Act interacts with the military domain, which reflects the difficulties encountered in relation to treating it as an exclusion. This iterative, consultative approach ensured that regulatory frameworks evolved through structured dialogue, which is a hallmark of effective CBMs.

The Act establishes clear regulatory, capability-based expectations that reduce uncertainty about how AI is governed and how guiding principles are to be interpreted. Its risk-based classification functions as a mechanism for foreseeability. When an AI application is still in its development stage, developers will already know what type of regulation they will need to adhere to once it reaches the market, while users will be able to use the regulations as a safety assurance.[75] At the same time, the risk-based approach articulates clear red lines for AI risks, which serve to signal and reiterate the normative priorities of the EU. Mandatory transparency and explainability standards also contribute to trust-building and accountability mechanisms in governance. By placing responsibility on developers to document their research, testing and validation processes, and to disclose AI-generated content, these standards help to reduce the uncertainty created by the black box nature of certain AI models.[76] Ultimately, these requirements reduce the likelihood that harmful systems will enter the market prematurely.[77] The European AI Board will function as a centralized body to reinforce regulatory compliance. This is especially important given how the Act elaborates the functions of human oversight. High-risk systems should always allow human intervention and people should bear ultimate responsibility. This specificity allows stakeholders to align shared norms with their understandings rather than fight over operationalization. This could have normative spillover effects beyond Europe as multinational companies align AI practices with EU regulations and the Act serves as a model for other AI governance frameworks.[78]

The potential for EU regulatory influence specifically in the realm of AI is currently more contested due to tensions between the EU and the USA. The latter continues to lead on innovation but the fact that the USA maintains a lead now does not make it inevitable that the EU will continue to lag behind. The European Commission announced a €200 billion AI investment initiative to bolster EU innovation in February 2025.[79] Furthermore, one of the key principles behind the €800 billion ReArm Europe Plan is to accelerate European development of AI and quantum technology, in a further demonstration of the EU's priorities and commitment to catching up.[80]

The EU AI Act demonstrates how structured regulatory processes, multi-stakeholder engagement and risk-based governance frameworks can function as capability- and behaviour-based CBMs. Undoubtedly, 'the EU has been a pioneer in holistic civilian AI governance' and has helped EU member states to move towards consensus.[81] Ultimately, however, the EU AI Act only 'underlines the urgent need to institutionalize stringent EU and international norms for military AI'.[82] Based on this initial analysis, section V explores how the content of and processes behind the drafting of the EU AI Act might provide lessons for the governance of AI in the military domain.

V. LESSONS LEARNED

Lessons for global governance deliberations

Prioritize the operationalization of principles

A key lesson on capability-based CBMs that can be drawn from the EU AI Act is the way in which it has operationalized key principles central to the AI governance debate. In many international forums, such as the GGE on LAWS and REAIM, most of the deliberations have centred on definitional debates.[83]

[74] European Centre for Not-for-Profit Law, 'EU AI Act needs clear safeguards for AI systems for military and national security purposes', 23 Mar. 2022; Csernatoni (note 11); and Dai and Song (note 56), p. 19.

[75] Siegmann and Anderljung (note 61).

[76] Hunter Christie, E. et al., 'Regulating Lethal Autonomous Weapon Systems: Exploring the challenges of explainability and traceability', *AI and Ethics*, 21 Feb. 2023; and Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S., 'Explainable AI: A review of machine learning interpretability methods', *Entropy*, vol. 23, no. 1 (2021), p. 18.

[77] Horowitz and Scharre (note 5), p. 7.

[78] Lingevicius (note 65), p. 3.

[79] European Commission, 'EU launches InvestAI initiative to mobilise €200 billion', Press release, 11 Feb. 2025.

[80] European Commission, 'Future of European Defence', [n.d.].

[81] Creutz et al. (note 54), p. 5; Csernatoni (note 11); and Dai and Song (note 56), p. 20.

[82] Csernatoni (note 11).

[83] Schmitt (note 9), p. 306; and Bode, I. et al., 'Prospects for the global governance of autonomous weapons: Comparing Chinese, Russian,

By choosing to draw red lines, the AI Act has become a conceptual reference point.[84] Seven focus areas emerged from the deliberations on military AI governance: (*a*) international law; (*b*) responsibility and accountability; (*c*) explainability and traceability; (*d*) bias and harm mitigation; (*e*) reliability; (*f*) governability; and (*g*) exchange of practices. In addition to its explicit emphasis on the exchange of practices, generally as a behaviour-based CBM, the EU AI Act also addresses the other six focus areas to varying degrees, potentially providing next steps for global deliberations.[85]

Accordance with international law. In addressing accordance with international law as a principle, the EU AI Act resolutely confirms that AI as a category of technologies will not be treated as an exception.[86] It not only acknowledges that AI is governed by human rights law and criminal law, as well as public international law, but also concretely underlines the responsibilities of states in these areas, including under international humanitarian law.[87] This stance addresses potential concerns about maintaining the EU's competitiveness, as compliance with international humanitarian law necessarily limits a state's choices of means and methods of warfare. Finally, the Act refers to the interaction of existing EU data and privacy regulation with uses of AI, further embedding this regulation in existing frameworks.[88]

In this way, the EU AI Act indicates three potential steps for the international governance deliberations around AI in the military domain. First, deliberations should make a concerted effort to identify which elements of international law apply to military AI and how, in order to both progress and concretize conversations within existing frameworks. Second, recognizing that international humanitarian law necessarily limits the options available to states in terms of tools of warfare could help to assuage some of the aspects of uncertainty outlined in section II, specifically those related to the ways in which AI has the potential to affect the character of war. Fundamentally, such thinking is already embedded in the stipulations of article 36 of Additional Protocol I (1977) to the 1949 Geneva Conventions, which imposes 'a practical obligation on states to review the legality of all new weapons, means or methods of warfare before they are used in an armed conflict'.[89] Finally, the EU AI Act stresses that its stipulations should not be addressed in isolation, but rather exist within a broader framework of regulation. The same is true of any emerging governance around AI in the military domain; it necessarily builds on and overlaps with existing technical standards and the regulation of related fields such as data and cyber.[90] This recognition helps not only to prevent duplication of efforts, but also to alleviate some of the regulatory uncertainty that stems from AI as a rapidly evolving technology. Regulation is not doomed always to lag behind.[91] These three lessons help to address the issues created by the lack of a governance end-game referred to in section I. If it is made clearer that the goal of emerging governance is not to be all-encompassing or to get it right from the start, such scoping could introduce the clarity required for progress.

Responsibility and accountability. The Act reaffirms that responsibility and accountability lie with actors throughout the AI lifecycle, with a specific obligation on developers and 'providers, distributor importers, deployers, third parties' in the private sector as they are at the forefront of innovation.[92]

An approach that clearly delineates the responsibilities and respective accountability measures of actors throughout the AI lifecycle could benefit deliberations around AI in the military domain. Certain responsibilities, such as those of the providers and distributors of AI technologies in the civilian domain, may have significant overlap with AI applications in the

and US practices', *Ethics and Information Technology*, vol. 25, no. 1 (Feb. 2023), p. 9.

[84] Dai and Song (note 56), p. 16.

[85] Sweijs and Romansky (note 3), p. 31.

[86] Tallberg, J. et al., 'The global governance of artificial intelligence: Next steps for empirical and normative research', arXiv, 19 May 2023, p. 11.

[87] Jurić (note 31), p. 4.

[88] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[89] Boulanin, V. and Verbruggen, M., 'SIPRI compendium on article 36 reviews', SIPRI Background Paper (Dec. 2017).

[90] Cristiano, F. et al. (eds), *Artificial Intelligence and International Conflict in Cyberspace* (Taylor & Francis, 2023).

[91] Bremmer, I. and Suleyman, M., 'The AI power paradox', *Foreign Affairs*, 16 Aug. 2023.

[92] Pacholska, M., 'Military artificial intelligence and the principle of distinction: A state responsibility perspective', *Israel Law Review*, vol. 56, no. 1 (Mar. 2023), p 5; and Regulation (EU) 2024/1689 of the European Parliament and of the Council (note 88).

military realm.[93] This creates a suitable starting point from which to impose any additional responsibilities on these actors and could allow deliberations around the military realm to focus on the responsibilities of domain-specific actors, such as military practitioners. This also addresses the dual-use nature of AI technologies and the uncertainty created by the involvement of multiple actors.

Explainability and traceability. To help actors to evaluate the risks posed by certain uses of AI, the principles of explainability and traceability are integrated into the Act as a prerequisite for accountability in the form of compliance verification and transparency.[94] While the Act's approach to explainability and traceability faces certain feasibility challenges due to the non-transparent nature of AI systems, these measures are still highly desirable. Explainability and traceability can help to accommodate the requirement to understand systems inputs and outputs, which is central to legal responsibility under international humanitarian law.[95]

Bias and harm mitigation. In the EU AI Act, a focus on explainability and traceability is often coupled with an emphasis on the importance of bias and harm mitigation. These measures are guided by the primacy of human rights and European values. For instance, article 45 stipulates 'non-discriminatory access to training data' while article 44c outlines safeguards to ensure bias detection.[96] As is discussed below, the way in which considerations about bias and harm mitigation are integrated into the EU AI Act reflects its focus on risks. For AI in the military domain as a whole, there is currently an opportunity to more prominently position human rights and dignity as the core guiding and unifying values for governance, and to recognize these as the foundation of international law.[97] This could also facilitate a risk-based approach, as governance would prioritize the category of risks to humans.

Reliability and governability. Finally, in the EU AI Act, reliability and governability are shaped as stringent requirements for human oversight to ensure maximum predictability and trust in systems. Article 15 and articles 49–51 emphasize that accuracy, robustness and (cyber)security in AI systems are a prerequisite for their deployment. Articles 62–68 set out how rigorous assessments of conformity with policy can make systems more trustworthy.[98] These elaborations enable the EU AI Act to set the normative boundaries within which norm development in the military domain should occur.[99]

As noted in section II, it is of paramount importance for militaries that any deployed systems are reliable, predictable and governable to the greatest extent possible. These notions are equally relevant in the civilian realm, especially in areas such as healthcare, medicine and public safety. In the military domain, the requirements are particularly stringent as they involve life-or-death decisions. It is therefore necessary to ensure that AI applications will function as intended during operations, and contribute to success rather than introduce risks of failure, excessive collateral damage or error.[100] Like the EU AI Act, ongoing governance initiatives could direct efforts towards identifying (in)appropriate consequences of AI use, including those that stem from emergent or unpredictable actions in AI systems.[101] Notably, the EU AI Act spotlights that these assessments must be integrated into the research and development stages of technologies, stressing that the reliability and trustworthiness of AI systems are dependent on proactive and forward-looking decision making.

By presenting a cohesive constellation of principles, the EU AI Act advocates for global norms that encourage responsible technological development and ethical uses of this emerging technology.[102] These principles are complemented by the EU's approach to the integration of military AI, as the AI Act and its other activities are primarily guided by the same

[93] Santoni de Sio, F. and Mecacci, G., 'Four responsibility gaps with artificial intelligence: Why they matter and how to address them', *Philosophy & Technology*, vol. 34, no. 4 (Dec. 2021), pp. 1057–84.

[94] Regulation (EU) 2024/1689 of the European Parliament and of the Council (note 88).

[95] Lewis, D. A., 'On "responsible AI" in war: Exploring preconditions for respecting International Law in armed conflict', eds S. Voeneky et al., *Cambridge Handbook of Responsible Artificial Intelligence* (Cambridge University Press: Cambridge, 2022), p. 500.

[96] Regulation (EU) 2024/1689 of the European Parliament and of the Council (note 88).

[97] Garcia, D., 'Algorithms and decision-making in military artificial intelligence', *Global Society*, vol. 38, no. 1 (Jan. 2024), p. 27.

[98] Regulation (EU) 2024/1689 of the European Parliament and of the Council (note 88).

[99] Sweijs and Romansky (note 3), p. 9.

[100] Pacholska (note 92).

[101] Trusilo, D., 'Autonomous AI systems in conflict: Emergent behavior and its impact on predictability and reliability', *Journal of Military Ethics*, vol. 22, no. 1 (Jan. 2023), p 5.

[102] Dai and Song (note 56), p. 21.

set of values. For instance, EDF regulations stipulate that 'any research project involving autonomous weapons should require meaningful human control', reiterating the EU's commitment to human oversight.[103] Values that are reflected in deliberation processes embed behaviour- and capability-based CBMs as mechanisms throughout the AI lifecycle, ensuring that all practices and policies on areas from risk management to documentation, auditability, robustness and cybersecurity help to foster knowledge of state-of-the-art solutions.[104] In the light of these observations, the governance of AI in the military domain can draw directly on how these principles have been operationalized, but also on the dynamics identified between them, building on the achievements of existing multilateral structures such as the EU.[105]

Direct focus on risk and risk-mitigation

Beyond the operationalization of values, norms and principles, the EU AI Act provides an approach to risk-mitigation through its tier system that could be employed in the governance of AI in the military domain. Global governance of military AI does not necessarily have to strive towards or adopt something as comprehensive as the AI Act. Nonetheless, it could serve as a model by emphasizing the importance of first obtaining a comprehensive overview of the elements that constitute risks and then finding methods for their mitigation.[106] No global governance initiative currently adopts a risk-based approach. This means that identification of risk is distributed throughout various discussions and academic publications, which does not necessarily contribute to structured deliberations.[107] Some initiatives have proposed a focused, risk-based approach to increase the likelihood of finding common ground on the categorization of which military systems should be prohibited, such as LAWS that facilitate war crimes, and which systems are permitted subject to the appropriate requirements on trustworthiness and accountability.[108] This could lead to an acknowledgement that in some cases the risks posed by the unpredictability of AI systems make certain uses illegal under international law, creating a suitable structure for regulating military AI or LAWS.[109]

The inherently high-risk nature of the military domain makes governance particularly challenging, as the technologies under discussion for multilateral oversight, such as autonomous weapons and AI-driven surveillance, are those with the greatest potential for strategic importance and those which raise the greatest ethical concerns. Despite the different risk thresholds and tolerances for risk, states are converging in certain areas, such as on the importance of strict rules on the integration of AI into nuclear command, control and communications, due to the significant impact of strategic miscalculations and the nuclear taboo.[110] Errors in interpretation, output or deployment in nuclear operations would be considerably more catastrophic than those in most other aspects of military activity and the nuclear taboo further heightens this particular concern. Identifying whether similar points of agreement can be found elsewhere, based on a clear overview of risks, in terms of potential consequences and respective probabilities, would be beneficial to global governance deliberations.

Continue with structured multi-stakeholder engagement

While most ongoing governance initiatives around AI in the military domain already practice multi-stakeholder engagement in various ways, the AI Act provides inspiration for additional modes of exchange that could contribute to CBMs, specifically on enhancing overall trust and transparency. Notably, the creation of the HLEG and the GTP as advisory bodies represented a deliberate and structured effort to engage with diverse perspectives. Collectively, these two groups formed a 'specialized knowledge network' to inform the EU bodies.[111] This approach aligns with the concept of 'front-door regulation', where private sector actors are formally included in decision-making processes through advisory panels, creating a structured and participatory model of governance.[112]

[103] Csernatoni and EUNPDC (note 12), p. 5.

[104] Siegmann and Anderljung (note 61); and Šonková, M., 'Brussels effect reloaded? The European Union's Digital Services Act and the Artificial Intelligence Act', EU Diplomacy Papers 4 and 5 (Mar. 2024).

[105] Ronnback (note 12).

[106] Cooper, Copeland and Sanders (note 33), p. 128.

[107] Arda, S., 'Taxonomy to regulation: A (geo)political taxonomy for AI risks and regulatory measures in the EU AI Act', arXiv, 17 Apr. 2024.

[108] Bruun, L., Bo, M. and Goussac, N., *Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What Does IHL Permit, Prohibit and Require?* (SIPRI: Stockholm, 2023); and Ronnback (note 12).

[109] Kwik and Van Engers (note 28). As most recently outlined in Blanchard, A. et al., 'A risk-based regulatory approach to Autonomous Weapon Systems', *Digital Society*, vol. 4, no. 1 (Apr. 2025), p. 23.

[110] Hass and Kahl (note 46).

[111] Dai and Song (note 56), p. 20.

[112] Bode, I. and Huelss, H., 'Constructing expertise: The front- and back-door regulation of AI's military applications in the European Union', *Journal of European Public Policy*, vol. 30, no. 7 (July 2023), p. 1235.

Table 3. Comparison of GC REAIM and RAISE

	Launch	Affiliated with	Membership	Representing	Focus	Outputs
Global Commission on Responsible AI in the Military Domain (GC REAIM)	2024	REAIM, founded by the Ministry of Foreign Affairs of the Netherlands and a number of international partners	18 Commissioners, around 40 experts, primarily from academia, as well as military and technical practitioners	**Global**, leaning towards European and North American membership	Linking dialogues between communities; supporting fundamental norm development and policy coherence	Expert Policy Notes (expected April–May 2025) Strategic Guidance Report (expected Sep. 2025)
Roundtable for AI, Security and Ethics (RAISE)	2024	The United Nations Institute for Disarmament Research (UNIDIR), Microsoft	21 members from government, industry and academia	**Global**	Acts as a catalyst for action, contributing to global governance based on cooperation, transparency, and mutual learning	Global Conference on AI, Security and Ethics (Mar. 2025)[a] Policy Brief (Sep. 2024)[b]

[a] UNIDIR, Global Conference on AI, Security and Ethics 2025, 27–28 Mar. 2025.

[b] Afina, Y. and Paoli, G. P., *Governance of Artificial Intelligence in the Military Domain: A Multi-Stakeholder Perspective on Priority Areas* (UNIDIR: Geneva, 2024).

However, private sector influence on AI governance extends beyond formal mechanisms, as 'back-door regulation' allows the technology itself to shape regulatory standards in ways that are less transparent and more subject to corporate interests.[113] This dual approach, which mixes formal state-led regulation with industry influence, results in a hybrid regulatory model that reflects the intrinsic role of private sector companies in developing dual-use AI technologies. The EU AI Act's consultative approach serves as a potential blueprint for integrating multi-stakeholder engagement into military AI governance.[114]

In current ongoing processes, only the creation of the Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM), as an offshoot of the REAIM process, and the Roundtable for AI, Security and Ethics (RAISE) as an initiative of the United Nations Institute for Disarmament Research (UNIDIR) in partnership with Microsoft approximate the EU's HLEG and GTP (see table 3). The UN Secretary-General's High Level Advisory Board on AI (HLAB) has focused on the security implications of AI technologies, but its priority is not to enhance governance.[115] GC REAIM has a dedicated multi-stakeholder mandate but does not involve industry or experts from key state actors. In 2024, GC REAIM also remained mostly separate from the REAIM process, which limited its immediate impact on deliberations.[116] Meanwhile, RAISE engages with a wider range of international experts and practitioners, and its independence from political processes provides a neutral and de-politicized space for discussion that can facilitate open exchanges without the constraints of formal diplomatic negotiations. Both GC REAIM and RAISE have the potential to play significant roles as sources of credible expertise and ideas. However, more structure and direction will be needed for the expert groups.[117]

Invest in the creation of an oversight body

Establishing a dedicated oversight body in the context of global governance around AI in the military domain could provide significant benefits in terms of the consolidation of capability-based CBMs, building on the precedent set by the EAIB. While legislation alone would be insufficient to ensure compliance and trust in AI governance, an independent, recognizable

[113] Bode and Huelss (note 112), p. 1236.

[114] Csernatoni and EUNPDC (note 12), p. 13; and Horowitz and Scharre (note 5).

[115] UNIDIR, 'RAISE: The Roundtable for AI, Security and Ethics', [n.d.].

[116] The Hague Centre for Strategic Studies, 'Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM)', 2024.

[117] Javadi, M. and Onderco, M., 'What does global military AI governance need?', European Leadership Network, Commentary, 2 Feb. 2024.

institution could function as a reliable source of expertise, assurance and norm diffusion. Such a body could be tasked with facilitating CBMs such as third-party audits, explainability standards and AI model testing frameworks, and help to evaluate the consistency of decision making, the quality of bias mitigation and the robustness of military AI use claims. Although it is unlikely that it would be able to enforce penalties for non-compliance with governance standards for AI in the military domain, a regulatory body could promote voluntary compliance frameworks and industry-led transparency initiatives that reinforce global AI governance efforts and maintain a commitment to human-centricity.[118]

In the short term, a body focused on expertise, assurance and norm diffusion is more likely to emerge, rather than one for auditing and standards, in the light of the current maturity of deliberations. A key actor such as UNIDIR could play a role in the formation of such a body, based on its global representativeness, credibility and involvement with processes such as REAIM and the GGE on LAWS. Specifically, UNIDIR's unique position as a neutral research institute within the UN ecosystem and its in-house expertise give it a potential role in supporting the eventual implementation of governance frameworks around AI in the military domain. For example, UNIDIR could help convene an International Expert Panel or Scientific Advisory Board, or continue to facilitate events like its inaugural 2025 Artificial Intelligence, Security and Ethics Conference. In this way, the institute could provide guidance, build on existing consensus and help to facilitate the operationalization of principles, a risk-based approach and multi-stakeholder engagement. While it is unlikely that UNIDIR would be able to play a comprehensive oversight role in and of itself, playing a role in the convening of information might still prove a valuable first step.

Some might argue that achieving consensus on military AI oversight is unrealistic. However, the EU has demonstrated that, despite its internal diversity, from its disparities in military capabilities to differing national security concerns, reaching agreement can be complex but not unachievable. The EU's outputs are also relevant as a supranational body directed to exercise some of the powers typically reserved to states, with established processes and means for doing so. Importantly, governance structures should not be rushed. The goal should be iterative development rather than achieving all the answers at once, allowing deliberations to adjust to technological and geopolitical shifts over time while building on and reiterating core values, whatever those may be determined to be.[119] The governance of AI in the military domain must be understood as a patchwork system that requires complementary governance efforts across multiple international forums rather than a single, overarching framework. At the current stage of development of global governance around AI in the military domain, the EU could serve as a moral compass at the intersection of normative and military power.[120]

Lessons for the EU and members states

Recognize a role in global governance deliberations for the EU and the small and middle European powers

The EU has the potential to become a valuable actor within global governance deliberations around AI in the military domain. First, the EU could gain a more secure sense of its 'right to act' if it strengthens its efforts to engage with member states, defence industries, international partners and global governance initiatives.[121] The EU currently participates in multilateral forums such as the UN Convention on Certain Conventional Weapons, the GGE on LAWS and REAIM, and cooperates with humanitarian organizations such as the International Committee of the Red Cross.[122] The legitimacy of such interactions is widely recognized: 73.2 per cent of stakeholders agree that 'the EU is willing to form a coalition for global military AI regulation' and 50.8 per cent believe it has the capacity 'to form [a] coalition with third countries and international organizations for the global regulation of military AI'.[123] At the same time, the EU member states represent a significant collection of global actors that are not only concerned with AI in the military domain, but have the potential to influence and continue to lead by example. European states such as Estonia, Finland, France, Germany and the Netherlands were among the first to raise the topic of AI in the military domain at the international

[118] Verbruggen and Boulanin (note 24).

[119] Lingevicius (note 65), p. 4.

[120] Lingevicius (note 65), p. 2.

[121] Csernatoni and EUNPDC (note 12), p. 12.

[122] Csernatoni and EUNPDC (note 12), p. 12; and Dai and Song (note 56), p. 20.

[123] Creutz et al. (note 54), p. 8.

level.[124] For example, the Netherlands was one of the founders of the REAIM summits and a main contributor to the First Committee resolution. This reflects the observation that small and middle powers in the EU are well positioned to play a particularly important role in global governance through CBMs.[125] For instance, small and middle powers played a leadership role in the negotiation of the 1997 Anti-Personnel Mine Ban Treaty (Ottawa Treaty) by leveraging diplomatic coalitions, moral authority and civil society partnerships to overcome resistance from major military powers. Canada, Norway, Austria and South Africa drove the process forward by emphasizing humanitarian concerns over security interests.[126]

An emboldened sense of actorness could help the EU to balance its normative leadership in responsible and ethical AI governance with the strategic realities of defence and security.[127] The EU AI Act leaves the bloc 'uniquely placed' to leverage its regulatory influence on civilian AI governance to extend it to military AI through multilateral engagement and coalition building.[128] By adopting a balanced framing that integrates normative power (human-centred AI regulation) and military power (pragmatic AI integration into defence), the EU can help to resolve some of the main challenges currently contributing to the deadlock in global governance deliberations.[129] For example, the EU member states and delegation helped to enshrine human control as the organizing principle of soft law instruments in the GGE on LAWS.[130] The EU has already called for global cooperation on regulating AI in the military domain but it must go beyond rhetoric by actively participating in these processes. This means that the EU must recognize the strategic and geopolitical dimensions of AI in the military domain beyond norm promotion. Experts hold that '[while] security and defense are not EU competencies, the Union cannot ignore the profound implications of the development and proliferation of military AI'.[131] This will mean leveraging both economic and diplomatic resources to forge flexible coalitions with strategic partners rather than relying solely on traditional multilateral institutions, and could ultimately help to ensure that internal military AI capabilities are backed by governance at multiple levels and beyond the EU.[132]

Work towards cohesion between the civilian and military domains

The consolidation of the EU's approach to AI regulation presents a unique opportunity to bridge the gap between governance of AI in the civilian and military domains. The norms, technical standards and governance principles developed for civilian AI through the EU AI Act's risk-based approach can serve as a foundation for shaping responsible, EU-wide policies on AI in the military domain.[133] At the same time, it is necessary to acknowledge that the potential impact of the EU in the international arena, as well as the benefits it can reap from global governance initiatives will be dependent on the cohesiveness of its internal policies. Misalignment could greatly detract from the EU's credibility as an actor by leaving 'political responsibility and risk management in the hands of Member States or, in the worst-case scenario, to the defence industry alone'.[134] Inconsistencies could also contribute to a risk of fragmentation where civilian regulatory frameworks promote the development of ethical AI while military AI is less stringently guided by ethical frameworks and less subject to external influence. This would present a significant obstacle if the EU wishes to retain control of AI regulatory development to prevent it from having to adopt standards set by other actors.[135] Finally, a cohesive framework would ensure that the full range of risks posed by the integration of AI into the military domain is addressed, while also making it possible to tap into the benefits of responsible uses of AI.[136]

The EU has historically followed a values-first approach to governance, prioritizing human-centricity, ethical safeguards and strong regulatory frameworks over setting binding legal standards.

[124] Jurić (note 31), p. 401.

[125] Creutz et al. (note 54), p. 16.

[126] Bolton, M. and Nash, T., 'The role of middle power-NGO coalitions in global policy: The case of the cluster munitions ban', *Global Policy*, vol. 1, no. 2 (May 2010), pp. 172–84.

[127] Creutz et al. (note 54), p. 17.

[128] Lingevicius (note 65), p. 11.

[129] Lingevicius (note 65), p. 13; Sweijs and Romansky (note 3), p. 8; and Schmitt (note 9).

[130] Specifically, France, Germany, Belgium, Spain, Italy, Latvia, the Netherlands, Poland, Bulgaria, Sweden, Finland, Ireland, Austria, Slovenia and the UK, which was still an EU member state at the time. Johansson-Nogués, Vlaskamp and Barbé (note 71), p. 23.

[131] Csernatoni (note 11).

[132] Schmid, S. et al., 'Arms race or innovation race? Geopolitical AI development', *Geopolitics*, 28 Jan. 2025, pp. 1–30.

[133] Csernatoni and EUNPDC (note 12), p. 13.

[134] Fanni, R., 'Why the EU must now tackle the risks posed by military AI', CEPS, 8 June 2023.

[135] Dai and Song (note 56), p. 19.

[136] Fanni (note 134).

This approach, seen in the successful implementation of the General Data Protection Regulation, has positioned the EU as a global leader in AI norm-setting and could similarly influence the governance of AI in the military domain.[137] However, the EU also faces internal limitations, as European states that favour total bans on LAWS have put in place self-imposed ethical and legal restraints on AI in the military domain in a way that often borders on 'cultural-technological conservatism'.[138] Given that EU external action operates predominantly within multilateral frameworks, the EU must expand its tech diplomacy and defence partnerships if it is to maintain strategic autonomy and competitiveness in military AI development while adhering to its ethical priorities.[139] This also means that the EU should take steps to ensure that its legal frameworks are not perceived as unnecessary and arduous red tape for industries and entities in the defence sector, but rather as advantageous in and of themselves.[140]

Bridging the gap between civilian and military AI governance is not a one-off regulatory task but an ongoing process.[141] As AI technologies evolve and geopolitical dynamics shift, the EU must continuously adapt its approach, refine its regulatory strategies and learn from international deliberations on AI governance. The EU's transnational nature inherently reflects broader global debates on AI governance, as member states exhibit divergent capabilities, interests and levels of technological development. By fostering cohesion between its civilian and military AI policies, the EU can strengthen its leadership on AI governance, enhance regulatory predictability and ensure that ethical considerations remain central to AI deployment in security and defence contexts.[142]

VI. CONCLUSIONS

Governance of AI in the military domain presents complex and evolving challenges. There are ongoing tensions between technological innovation and security priorities, as well as ethical and legal considerations. The emergence of multiple global governance initiatives demonstrates a growing consensus on foundational principles and an urgent need to address the risks arising from the integration of AI into the military domain. Nonetheless, a gap remains between high-level principles and practical regulation, which has contributed to regulatory stagnation. CBMs offer a crucial path forward, providing mechanisms to enhance trust, reduce uncertainty and promote structured dialogue among states.

The EU AI Act offers lessons for both capability- and behaviour-based CBMs for AI in the military domain as a dual-use technology. The content of and processes behind the AI Act underline the utility of a risk-based framework, multi-stakeholder engagement and oversight mechanisms. By integrating these lessons into global governance deliberations, high-level principles can be operationalized in a structured and enforceable manner. At the same time, the EU AI Act also highlights that the EU and its member states must strengthen their engagement with global military AI governance, as 'technology is key to both self-perceptions and international perceptions of the EU's status as one of the leading geopolitical players'.[143]

An effective AI governance strategy must therefore ensure cohesion between civilian and military regulatory frameworks in order to leverage normative influence. Ultimately, the challenge the EU faces in the global governance of military AI goes beyond a simple regulatory one, as there are vital foreign policy, moral and strategic priorities and imperatives at play.[144] By aligning norm promotion with coalition building, the EU can reinforce its global governance role while safeguarding its long-term security interests. At the same time, by treating the work of the EU as a reference point, international efforts can shift from merely rhetorical commitments to coherent, risk-informed and practical frameworks for governance of the responsible development, deployment and use of AI in the military domain.

[137] Šonková (note 104).
[138] Jurić (note 31), p. 428; see also Soare (note 10); and Greene (note 47).
[139] Soare (note 10), p. 87.
[140] Jurić (note 31), p. 428.
[141] Mügge, D., 'Regulatory interdependence in AI', eds R. Paul, E. Carmel and J. Cobbe, *Handbook on Public Policy and Artificial Intelligence* (Edward Elgar Publishing: Cheltenham, 2024), pp. 249–60.
[142] Lingevicius (note 65), p. 4.
[143] Soare (note 10), p. 77.
[144] Csernatoni (note 11).

This document has been produced with the financial assistance of the EU. The contents are the sole responsibility of the EU Non-Proliferation and Disarmament Consortium and can under no circumstances be regarded as reflecting the position of the EU.

A EUROPEAN NETWORK

In July 2010 the Council of the European Union decided to support the creation of a network bringing together foreign policy institutions and research centers from across the EU to encourage political and security-related dialogue and the long-term discussion of measures to combat the proliferation of weapons of mass destruction (WMD) and their delivery systems. The Council of the European Union entrusted the technical implementation of this Decision to the EU Non-Proliferation Consortium. In 2018, in line with the recommendations formulated by the European Parliament the names and the mandate of the network and the Consortium have been adjusted to include the word 'disarmament'.

STRUCTURE

The EU Non-Proliferation and Disarmament Consortium is managed jointly by six institutes: La Fondation pour la recherche stratégique (FRS), the Peace Research Institute Frankfurt (HSFK/ PRIF), the International Affairs Institute in Rome (IAI), the International Institute for Strategic Studies (IISS–Europe), the Stockholm International Peace Research Institute (SIPRI) and the Vienna Center for Disarmament and Non-Proliferation (VCDNP). The Consortium, originally comprised of four institutes, began its work in January 2011 and forms the core of a wider network of European non-proliferation and disarmament think tanks and research centers which are closely associated with the activities of the Consortium.

MISSION

The main aim of the network of independent non-proliferation and disarmament think tanks is to encourage discussion of measures to combat the proliferation of weapons of mass destruction and their delivery systems within civil society, particularly among experts, researchers and academics in the EU and third countries. The scope of activities shall also cover issues related to conventional weapons, including small arms and light weapons (SALW).

www.nonproliferation.eu

EU Non-Proliferation and Disarmament Consortium

Promoting the European network of independent non-proliferation and disarmament think tanks

FONDATION pour la RECHERCHE STRATÉGIQUE

FOUNDATION FOR STRATEGIC RESEARCH

www.frstrategie.org

PRIF Peace Research Institute Frankfurt HSFK Hessische Stiftung Friedens- und Konfliktforschung

PEACE RESEARCH INSTITUTE FRANKFURT

www.hsfk.de

iai Istituto Affari Internazionali

INTERNATIONAL AFFAIRS INSTITUTE

www.iai.it/en

IISS

INTERNATIONAL INSTITUTE FOR STRATEGIC STUDIES

www.iiss.org/en/iiss-europe

sipri

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

www.sipri.org

VCDNP Vienna Center for Disarmament and Non-Proliferation

VIENNA CENTER FOR DISARMAMENT AND NON-PROLIFERATION

www.vcdnp.org