The Vienna Document 2011 and Military Applications of Artificial Intelligence

Nicolò Miotto[*]

Abstract

The development and deployment of military applications of artificial intelligence (AI) is raising concerns about their negative implications for international security. Misperception, unintended escalation, and proliferation are some of the key potential risks stemming from military uses of AI. This article argues that states within and outside the OSCE region should draw on the OSCE Vienna Document 2011 to develop confidence- and security-building measures (CSBMs) applicable to the military uses of AI. Such CSBMs could help foster dialogue and co-operation by increasing transparency and predictability concerning military applications of AI.

Keywords

OSCE, artificial intelligence, Vienna Document 2011, CSBMs, military transparency

To cite this publication: Nicolò Miotto, "The Vienna Document 2011 and Military Applications of Artificial Intelligence," in *OSCE Insights*, eds. Cornelius Friesendorf and Argyro Kartsonaki (Baden-Baden: Nomos, 2025), https://doi.org/10.5771/9783748945857-01

Introduction

Artificial intelligence (AI) is expected to bring about unprecedented innovation in numerous sectors of society, including defense.[1] Its use in the military promises various technical benefits, including improvements in data collection, strengthened analytical capabilities, and faster decision-making processes. As several countries have manifested their interest in developing military applications of AI, a fierce public debate surrounding their potential technical, (geo)political, and ethical risks has been taking place. While some observers have highlighted that, despite the risks, AI can improve key military capabilities such as early warning and target identification, others have warned against potential risks such as misperception, unintended escalation, and proliferation.[2] In noting these challenges, many have engaged in reflection on potential means of mitigating such threats.

Among other tools, diverse stakeholders have suggested developing confidence- and security-building measures

* OSCE Conflict Prevention Centre
nicolo.miotto@osce.org;
nicolo.miotto@libero.it

(CSBMs) for military applications of AI to increase transparency, enhance predictability, and avert escalation. Hence, research on CSBMs is expanding, receiving contributions from academia, governments, and the private sector.[3] With that said, these studies mainly focus on developing new measures that can address both the technical limitations of AI and their potential implications for international security. Little attention has been paid to exploring the applicability of well-established CSBMs to the military uses of AI. In particular, what is lacking—with the single exception of a rather general study[4]—is an analysis of the contribution that the OSCE Vienna Document 2011 (VD11) could make in this regard.[5]

Reflecting on the contributions of the VD11 to the multilateral governance of military uses of AI is of the utmost importance at a time when international discussions on the matter have stalled.[6] Due to the erosion of trust and confidence caused by Russia's war of aggression against Ukraine, it is unlikely that the VD11 will be updated any time soon to cover military applications of AI. Nonetheless, this study argues that states within and outside the OSCE region should draw upon the VD11 to implement CSBMs to increase the transparency and predictability of military uses of AI.

This paper starts by outlining the definitions of AI and CSBMs adopted in this research. It then addresses prominent issues pertaining to military uses of AI and key CSBMs that have been recommended to mitigate related threats. It then explores the main problems underlying the application of CSBMs to military uses of AI, noting that despite these challenges, certain arrangements could likely be implemented successfully. Finally, it shows how key VD11 provisions could be drawn on to establish CSBMs for military uses of AI and provides recommendations in this direction.

Definitions and terminology

Artificial intelligence and its military applications

AI is a much-used umbrella concept that incorporates numerous related technologies and areas of research, including machine learning (ML) and deep learning (DL). Definitions of AI vary depending on the capabilities of the systems in question and their functionalities.[7] Despite their diversity, however, these definitions point to certain general features related to the overall rationale and objectives of AI technologies. Such characteristics include the capacity to simulate human reasoning and perform cognitive tasks that are generally associated with human intelligence.[8]

A closer look at the quantity and quality of the cognitive tasks simulated by these technologies helps to further clarify what AI is by marking the difference between so-called "artificial general intelligence" (AGI)/"artificial super intelligence" (ASI) and "narrow AI." AGI/ASI represents a strictly hypothetical form of AI which would be capable of equaling or surpassing human intelligence and behavior, becoming self-conscious and

acquiring the ability to perform tasks, learn, and plan autonomously as humans do.[9] The category of narrow AI, to which current uses of AI belong, comprises "complex software programs that can execute discrete 'intelligent' tasks such as recognizing objects or people from images, translating language, or playing games."[10] Narrow AI programs include ML and its sub-field, DL.

This paper looks at military applications of AI as an ensemble of narrow AI programs used to carry out specific military tasks such as image recognition, autonomous navigation, and training. This research only considers uses of narrow AI to enhance the capabilities of the weapon and equipment systems covered by the VD11 (e.g., battle tanks, armored combat vehicles, and combat aircrafts).[11] Therefore, certain conventional and non-conventional weapon and equipment systems not covered by the VD11, such as warships and nuclear command, control, and communications, are not considered by this study.

Confidence- and security-building measures (CSBMs)

This paper adopts a general definition of CSBMs, as outlined in early research, as arrangements designed to enhance

> an assurance of mind and belief in the trustworthiness of the announced intentions of other states in respect of their security policies, and the facts with regard to military activities and capacities which are designed to further the objectives of a nation's security policy.[12]

The main objectives of CSBMs are to increase transparency by publicly displaying a state's non-aggressive posture and to enhance predictability by allowing for the detection of inconsistencies in other states' behavior vis-à-vis established CSBMs.[13] The ultimate intended impact of CSBMs is to reduce the risk of unintended escalation and conflict between countries, which could be triggered by misperceptions about other states' military postures and activities. Examples of CSBMs include the notification of military exercises, the observation of military activities, the establishment of communication channels between countries, inspections of military facilities, and the exchange of information on military forces and budgets.[14] These cases mirror the principles and practices outlined in pivotal OSCE documents such as the 1975 Helsinki Final Act[15] and the VD11.

Military applications of AI, associated risks, and CSBMs

Several countries, including the United States, Russia, and China, are heavily investing in AI to modernize their military capabilities.[16] This interest in developing military applications of AI stems from the technical opportunities they offer (such as improvements in target identification and the acceleration of decision-making processes)[17] and from the ambition to equal or surpass competitors' actual and/or perceived capabilities.[18] Projects

aimed at integrating AI into military systems encompass a wide range of tools, including unmanned aerial and maritime vehicles, missile technology, nuclear capabilities, and space systems. AI is being developed and tested to support other military tasks, including command and control, information management, logistics, and training.[19] Existing AI capabilities in these sectors include collateral damage estimation, the geolocation of images, the provision of recommendations on best paths and transport modes, and the tracking of individuals' learning progress.[20] The strong interest in further improving these tools and developing new ones is driven by the advantages AI offers, such as enhanced assessment accuracy, faster analysis and communication, and lower logistics costs.[21]

Despite these promising opportunities, researchers, public institutions, and civil society organizations have expressed several concerns about the military uses of AI. Indeed, the technology is vulnerable to several limitations. For instance, technical issues such as changes in the data distribution can negatively impact the performance of AI models.[22] Furthermore, malicious actors can affect the integrity of data by manipulating the training datasets, thus leading AI models to fail or to act differently than expected.[23] Additional issues such as psychological constraints can affect human-machine interactions; for example, end-users can act upon erroneous analytical outputs due to unconditional trust in AI data analysis capabilities.[24]

In a military context, these and further issues can have serious security implications, potentially undermining international security. Possible technical failures range from errors in autonomous navigation to target misidentification, paving the way for concerning scenarios such as diplomatic tensions, escalation, and even overt military conflict.[25] In response to these challenges, academics, policymakers, and private companies have recommended different types of CSBMs. These can be grouped into two main categories based on the issues they aim to address: (1) CSBMs that address potential technical issues with AI software; and (2) CSBMs that address inter-state security dynamics underlying the development and deployment of military applications of AI. The first category includes measures such as the publication of system cards[26] to provide information about the capabilities and limitations of AI models and the use of content provenance and watermarking methods to verify the authenticity and integrity of AI-generated data.[27]

CSBMs from the second category include broader arrangements such as the establishment of Track II initiatives[28] to promote dialogue on the risks posed by military uses of AI and the releasing of joint political declarations on the maintenance of human control over decisions concerning target engagement.[29] Additional measures include tabletop exercises to simulate crisis scenarios and develop tailored responses, the establishment of hotlines between countries, and the development of incident sharing agreements to consolidate knowledge of AI technical failures and their impact on security.[30]

These CSBMs represent valuable measures to mitigate key potential threats. However, their effective implementation faces several challenges stemming from the current geopolitical environment and the intrinsic characteristics of AI technology. Analyzing these limitations can help us to understand which CSBMs are more likely to contribute to the goals of enhancing transparency and predictability.

Challenges and opportunities for the application of CSBMs to the military uses of AI

Geopolitical and technical challenges

While the need to engage in talks about military applications of AI and their regulation has been recognized by the academic and policymaking community, several dilemmas continue to pose obstacles to the implementation of concrete measures. Geopolitical tensions following Russia's war of aggression against Ukraine represent a prominent example of the challenges affecting the negotiation of CSBMs. Indeed, CSBMs can be seen as the ultimate representation of a shared understanding of what constitutes common security concerns.[31] Their effective negotiation depends on the establishment of confidence and trust between states. Hence, their development is conditional on rebuilding trust and confidence and achieving a common notion of which issues pertaining to military applications of AI represent security matters of reciprocal interest.

Moreover, in such a contested environment, it is unlikely that states will adopt intrusive AI software–focused CSBMs such as system cards. This has already been highlighted in the research on cyber CSBMs, which notes that non-like-minded countries are unlikely to implement intrusive measures such as the observation of cyber exercises in order to maintain a degree of secrecy over cyber capabilities.[32] Indeed, states that have deployed cutting-edge military applications of AI are unlikely to publicly acknowledge the limitations or potential biases that affect their functioning, especially vis-à-vis adversaries' deployment of such technologies. This would be detrimental to their security interests and could reveal gaps in military effectiveness. When AI software transparency is weighed up against the projection of military power, the balance often tips in favor of the latter.

Dilemmas inherent to the technology only add to these geopolitical challenges. As noted by recent research, there is much uncertainty about whether AI and its military applications can be effectively tested to verify that systems are functioning and behaving as originally intended, designed, and expected and about which techniques and methods can be employed to best conduct technical assessments.[33] This overall uncertainty has serious implications for CSBMs as it calls into doubt what can be verified with certainty about the military uses of AI. In the face of this uncertainty, not only are countries likely to refrain from implementing AI software–related CSBMs, but, even if circumstances were different, they

would face technical challenges to effectively ensuring the safety of military uses of AI.

Despite these notable challenges, shedding light on existing co-operative dynamics between states in the international environment and shifting the focus from AI software to military hardware can help us to assess whether less intrusive measures are more feasible and can be effectively implemented.

Opportunities for politically and technically feasible CSBMs

While the security environment is competitive and characterized by strong tensions, multilateral discussions on the military applications of AI have already taken place at intergovernmental fora before and following Russia's war of aggression against Ukraine, including at the OSCE. At the OSCE, formal and informal discussions have been particularly focused on the impact of AI on law enforcement and crime,[34] freedom of expression and media pluralism,[35] human rights,[36] and international law.[37] Attention has also been paid to the military uses of AI. For example, informal discussions on these issues took place between 2014 and 2021, bringing to the table governmental and non-governmental representatives from OSCE participating States.[38]

Most importantly, from 2019 to 2021 the OSCE Parliamentary Assembly (PA) and the Forum for Security Co-operation (FSC) hosted formal political discussions between OSCE participating States on the military uses of AI.[39] Such engagement also included discussions on whether existing arms control frameworks, including the VD11, should be updated to account for the military uses of AI. While such discussions have not taken place at either the PA or the FSC recently, they have continued in other formats, expanding formal political engagement beyond Europe by including the OSCE Asian Partners for Co-operation.[40]

Therefore, while geopolitical tensions are hindering in-depth discussions on the overall arms control architecture and eroding trust and confidence, evidence also points to the fact that more limited but important informal and formal discussions are already taking place at the multilateral level within and outside the OSCE region. Although such engagement primarily involves like-minded countries, it nevertheless represents an important step, paving the way for future discussions when the security environment allows.

Technical issues concerning the verification and validation of AI software should not overshadow the potential benefits of applying less intrusive and more technically feasible CSBMs to AI-integrated military hardware.[41] Research on cyber CSBMs has shown that arrangements such as the exchange of information on cyber doctrines and the organization of cyber forces are likely to be implemented, even among non-like-minded countries.[42] Moreover, likeminded states are more open to discussing and implementing even intrusive CSBMs such as those concerning the prior notification and observation of military cyber

exercises.[43] This is not mere theory, as the OSCE already represents an existing successful model. Between 2013 and 2016, the Organization served as a platform for adopting a total of sixteen voluntary cyber CBMs which encompass a wide set of arrangements, ranging from information exchanges on cyber doctrines, strategies, and policies to the voluntary reporting of cyber vulnerabilities.[44]

Furthermore, key CSBMs can be applied to AI-integrated military hardware. For example, if a state were to deploy an unmanned aerial vehicle (UAV) equipped with AI autonomous navigation software to better conduct military intelligence gathering at its borders, its neighbors may be more interested in why it deployed such technology and whether this indicates a change in its military posture than in whether the UAV's AI software works effectively. This observation opens the door for the implementation of certain CSBMs to increase transparency between states by signaling a non-aggressive military posture and to enhance predictability by helping to detect anomalies in states' behavior. If the AI software cannot be inspected due to security concerns, secrecy requirements, and lack of effective methodologies, then measures should focus on the deployment of military hardware and its implications. In this sense, the VD11 could serve as a basis for implementing concrete measures to mitigate certain detrimental inter-state security dynamics underlying the development and deployment of military applications of AI.

CSBMs for military uses of AI: The VD11 as a source

The VD11 does not cover military uses of AI, and therefore its applicability to this domain is strictly dependent on future updates to the document. Due to existing politico-military tensions, it is unlikely that the VD11 will be amended in the near future. Nonetheless, OSCE participating States should draw upon VD11 provisions to create voluntary CSBMs to increase transparency and predictability concerning the military uses of AI. Similarly, states outside the OSCE region should use the VD11 as an inspiration for similar measures. The feasibility of applying the various CSBMs outlined in VD11 to military uses of AI can be assessed following the same logic as that used in the previous section's discussion of which measures are more likely to be implemented in the near future. The CSBMs set out in the VD11 offer a crucial means of improving transparency, allowing states to assess each other's intentions and military postures. They could also enhance predictability by providing diplomatic channels for discussing states' behavior with regard to the development and employment of military applications of AI.

Because it is unlikely that states will adopt intrusive CSBMs allowing for the inspection of AI software, other more feasible VD11 arrangements could be considered. Moreover, because it is highly difficult to validate and verify AI models,[45] such arrangements would need to tackle other issues first. For example, states could address the destabilizing

implications of reciprocal uncertainty concerning military budget allocations and weapons development.[46] Additionally, countries could dispel concerns related to newly developed military doctrines that contemplate the use of new and emerging technologies.[47] If they are not addressed, these matters risk destabilizing inter-state relations, leading to misperceptions and erroneous assessments of other countries' intentions and military postures. These uncertainties are particularly impactful in the case of AI since states are competing to develop its military applications and, consequently, are heavily investing in this endeavor.[48] The VD11 contains numerous CSBMs to shed light on military expenditure, military research and development, and military doctrines and strategies, thus providing an effective means of assessing countries' intentions.

While it is unlikely that states will implement CSBMs concerning the demonstration of military cyber capabilities,[49] this does not necessarily apply to the military uses of AI. Indeed, if the capabilities are looked at from a hardware (rather than a software) perspective, states may be interested in showcasing how AI is being employed to enhance the performance of a given weapon and equipment system. For instance, a state might be interested in demonstrating (including to its adversaries) its use of AI to improve the navigation capabilities of an armored vehicle, as a means of showcasing advances in its defense capabilities. In doing so, it would not need to share the technical characteristics of the AI software, the algorithm underlying the ML model, or the training dataset used. Certainly, such a demonstration would be limited in scope, but it would provide insight into how that state intends to use military applications of AI. The VD11 therefore offers an important basis for providing general information about AI-integrated weapon and equipment systems.

Although intrusive CSBMs are less likely to be implemented, this does not mean that arrangements should not consider the security implications of potential technical failures of AI software. Indeed, a mere technical failure could be read as a discrepancy in a state's behavior and military posture and could thus generate tensions. If the autonomous navigation system of an AI-powered UAV were to fail, for example, causing it to accidentally cruise into the airspace of a rival neighboring country, this could be mistakenly interpreted as a hostile act. In such cases, there is a need to quickly reassure adversaries in order to dispel concerns and avert unintended escalation. In this sense, crisis hotlines are a valuable means of responding to such emergencies. The VD11 provides for well-structured measures that could support states under these circumstances.

Recommendations

The following recommendations focus on often overlooked but prominent VD11 CSBMs, in particular key provisions outlined in Chapter II ("Defence Planning"), Chapter III ("Risk Reduction"), and Chapter IV ("Contacts"). These measures, in contrast to provisions such as the annual exchange of military

information, have yet to receive sufficient attention. In addition, they provide a feasible field for action in contrast to other VD11 provisions such as Chapter VI ("Observation of Certain Military Activities"), which would likely be perceived as particularly sensitive and intrusive. Drawing on the CSBMs set out in the VD11, states within and outside the OSCE region should consider:

Implementing information exchange on defense planning concerning military applications of AI. VD11 Chapter II, "Defence Planning," foresees information exchange between OSCE participating States regarding their

> intentions in the medium to long term as regards size, structure, training and equipment of [their] armed forces, as well as defence policy, doctrines and budgets related thereto.[50]

The exchange of such information aims to increase transparency and promote dialogue between participating States. These provisions require participating States to exchange information on the "training programmes for their armed forces and planned changes thereto in the forthcoming years," as well as the "procurement of major equipment and major military construction programmes […], either ongoing or starting in the forthcoming years."[51] In addition, if information is available, participating States are expected to provide "the best estimates specifying the total and figures for […] research and development" with regard to the last two years of the forthcoming five fiscal years.[52] As part of their information exchange, OSCE participating States should consider the voluntary provision of details and estimates on budget allocations, military research and development, AI-integrated weapon and equipment systems, and new military doctrines that include the employment of military applications of AI. States outside the OSCE region should establish similar mechanisms to provide insights into their intensions and military postures in the medium and long term.

Using existing platforms and/or developing new ones to discuss the information exchanged. According to VD11 Chapter II, any participating State can ask for clarification on the defense planning–related information provided by another participating State. High-level discussions on the information are envisaged in the format of the Annual Implementation Assessment Meeting (AIAM), the High-Level Military Doctrine Seminar (HLMDS), and study visits.[53] The HLMDS is a particularly relevant format for discussing such matters. It brings together high-level military and civilian representatives such as chiefs of defense and/or chiefs of general staff, diplomats, and academics, who discuss doctrinal changes, their impact on military structures, and the military information exchanged. OSCE participating States should consider voluntarily discussing the information exchanged at the HLMDS. States outside the OSCE region should use similar structures or develop new ones to engage in dialogue on the impact of AI on military structures and doctrines, exchanging views on white papers, defense policies, and military doctrines.

Establishing co-operation as regards hazardous incidents of a military nature involving military applications of AI. VD11 Chapter III.17, "Co-operation as Regards Hazardous Incidents of a Military Nature," outlines measures to prevent possible misunderstandings in the event of a military incident.[54] If a hazardous incident of a military nature occurs, the participating State whose military forces are involved in the incident should provide information to other participating States, and any participating State affected by the incident can also request clarification. This general mechanism could be employed in the event of incidents involving military applications of AI such as the hypothetical cases concerning AI-powered UAVs outlined in the previous sections. In line with the provisions of this chapter, participating States have an established point of contact (PoC) to better co-ordinate communications in the event of a hazardous incident of a military nature. In the context of military uses of AI, participating States should employ this mechanism to dispel concerns. States outside the OSCE region should develop similar measures, such as crisis hotlines, thus reducing the risk of accidental military escalation. PoCs can quickly provide both technical and political information to the relevant counterpart(s), warning against potential weapon system failures and dispelling concerns about the nature of the military activity.

Holding discussions on hazardous incidents of a military nature involving military applications of AI. As outlined in Chapter III.17, hazardous incidents of a military nature can be discussed at the FSC and at the AIAM.[55] In the context of the military applications of AI, these discussions could help to clarify the nature of the incidents and to pave the way for broader dialogue on the security risks posed by AI and means of averting escalation. In particular, discussions could address the possible repercussions of diverse technical malfunctions for international security. OSCE participating States should hold these talks at the AIAM to foster dialogue. States outside the OSCE region should bring discussions to existing venues or create new platforms for discussing such matters.

Using existing data-sharing tools and/or developing new ones as incident sharing repositories. Details on incidents involving military uses of AI such as location, type of weapon or equipment system involved, and the nature of the incident (for example airspace infringement, target misidentification) should be shared between states within and outside the OSCE region. An example of a data-sharing tool that participating States could employ is the OSCE Communications Network, which is used for information exchange under the VD11. Following the example of the Communications Network, states outside the OSCE region should develop data-sharing tools to share information on the incidents and engage in political discussions informed by accurate, evidence-based analyses.

Organizing demonstrations of new types of AI-integrated major weapon and equipment systems. VD11 Chapter IV.31, "Demonstration of New Types of Major Weapon and Equipment Systems," requires any participating State that

deploys "a new type of major weapon and equipment system" to "arrange […] a demonstration for representatives of all other participating States."[56] As countries are deploying military applications of AI, these demonstrations could be particularly helpful in creating occasions for dialogue and co-operation. Participating States should consider applying this CSBM to the military uses of AI. Accordingly, participating States that deploy new types of AI-integrated major weapon and equipment systems should arrange demonstrations for the representatives of all other participating States. For instance, a participating State could demonstrate how new types of armored vehicles employ autonomous navigation for path planning and real-time path adjustment and explain how these new types of weapon and equipment systems fill the gaps of previous versions of military hardware. States outside the OSCE region should consider implementing similar measures at the bilateral and multilateral levels. Notably, such demonstrations would still allow countries to maintain their technological advantage, as general information about the relevant military hardware capabilities could be shared without requiring the sharing of AI software.

Discussing the results of the demonstrations. According to VD11 provisions, following up on the demonstrations, participating States can discuss observations and results at key OSCE fora such as the FSC and the AIAM. States outside the OSCE region should bring these discussions to existing regional fora or develop new venues for such engagement. Such discussions could be particularly valuable as opportunities not only for addressing present concerns but also for raising technical and political matters related to future deployments of military applications of AI.[57]

Notes

1 Darrell M. West and John R. Allen, *Turning Point: Policymaking in the Era of Artificial Intelligence* (Washington: Brookings Institution Press, 2020).

2 See Jessica Cox and Heather Williams, "The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability," *Washington Quarterly* 44, no. 1 (2021): 69–85; István Szabadföldi, "Artificial Intelligence in Military Application: Opportunities and Challenges," *Land Forces Academy Review* 26, no. 2 (2021): 157–65.

3 Michael C. Horowitz and Paul Scharre, AI and International Stability: Risks and Confidence-Building Measures (Washington, DC: Center for a New American Security, 2021), https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/AI-and-International-Stability-Risks-and-Confidence-Building-Measures.pdf; Marina Favaro, Strengthening the OSCE's Role in Strategic Stability (Atlantic Council, 2022), https://www.atlanticcouncil.org/wp-content/uploads/2022/01/Strategic-Insights-Memo_OSCE-and-Strategic-Stability_1.12.22-1.pdf; Anna Nadibaidze, Commitment to Control Weaponised Artificial Intelligence: A Step Forward for the OSCE and European Security (Geneva: Geneva Centre for Security Policy, 2022), https://www.gcsp.ch/publications/commitment-control-weaponised-artificial-intelligence-step-forward-osce-and-european; Sarah Shoker et al., "Confidence-Building Measures

for Artificial Intelligence: Workshop Proceedings," arXiv:2308.00862 [cs.CY], arXiv, August 3, 2023, https://arxiv.org/abs/2308.00862

4 Favaro, cited above (Note 3).

5 OSCE, Vienna Document 2011 on Confidence- and Security-Building Measures, FSC.DOC/1/11 (Vienna: November 30, 2011), https://www.osce.org/fsc/86597

6 Ingvild Bode et al., "Prospects for the Global Governance of Autonomous Weapons: Comparing Chinese, Russian, and US Practices," *Ethics and Information Technology* 25, no. 1 (2023): Article 5.

7 IBM Data and AI Team, "Understanding the Different Types of Artificial Intelligence," IBM, October 12, 2023, https://www.ibm.com/blog/understanding-the-different-types-of-artificial-intelligence/

8 Ralf T. Kreutzer and Marie Sirrenberg, "What Is Artificial Intelligence and How to Exploit It?," in *Understanding Artificial Intelligence: Fundamentals, Use Cases and Methods for a Corporate AI Journey* (Cham: Springer, 2020), 1–57.

9 Scott McLean et al., "The Risks Associated with Artificial General Intelligence: A Systematic Review," *Journal of Experimental & Theoretical Artificial Intelligence* 35, no. 5 (2021): 649–63.

10 Vincent Boulanin, ed., *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. 1, *Euro-Atlantic Perspectives* (Stockholm: SIPRI, 2019), 14, https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf

11 The full list of weapon and equipment systems is reported in Annex III of the VD11.

12 Johan Jørgen Holst, "Confidence-Building Measures: A Conceptual Framework," *Survival* 25, no. 1 (1983): 2.

13 Abbott A. Brayton, "Confidence-Building Measures in European Security," *The World Today* 36, no. 10 (1980): 382–91; Erica D. Borghard and Shawn W. Lonergan, "Confidence Building Measures for the Cyber Domain," *Strategic Studies Quarterly* 12, no. 3 (2018): 10–49.

14 Holst, cited above (Note 12); Brayton, cited above (Note 13).

15 CSCE, Helsinki Final Act (Helsinki: 1975), https://www.osce.org/helsinki-final-act

16 Margarita Konaev et al., U.S. Military Investments in Autonomy and AI: A Strategic Assessment (Center for Security and Emerging Technology, 2020), https://cset.georgetown.edu/wp-content/uploads/U.S.-Military-Investments-in-Autonomy-and-AI_Strategic-Assessment-1.pdf; Samuel Bendett et al., Advanced Military Technology in Russia: Capabilities and Implications (London: Chatham House, 2021), https://www.chathamhouse.org/sites/default/files/2021-09/2021-09-23-advanced-military-technology-in-russia-bendett-et-al.pdf

17 Eric Robinson, Daniel Egel, and George Bailey, Machine Learning for Operational Decisionmaking in Competition and Conflict: A Demonstration Using the Conflict in Eastern Ukraine (Santa Monica, CA: RAND Corporation, 2023), https://www.rand.org/pubs/research_reports/RRA815-1.html

18 Anna Nadibaidze and Nicolò Miotto, "The Impact of AI on Strategic Stability Is What States Make of It: Comparing US and Russian Discourses," *Journal for Peace and Nuclear Disarmament* 6, no. 1 (2023): 47–67.

19 Elsa B. Kania, "Chinese Military Innovation in the AI Revolution," *RUSI Journal* 164, no. 5–6 (2019): 26–34; Thomas Reinhold and Niklas Schörnig, eds., *Armament, Arms Control and Artificial Intelligence: The Janus-Faced Nature of Machine Learning in the Military Realm* (Cham: Springer Nature, 2022); Sarah Grand-Clément, Artificial Intelligence beyond Weapons: Application and Impact of

AI in the Military Domain (Geneva: UNIDIR, 2023), https://unidir.org/wp-content/uploads/2023/10/UNIDIR_AI_Beyond_Weapons_Application_Impact_AI_in_the_Military_Domain.pdf

20 Kania, cited above (Note 19); Reinhold and Schörnig, cited above (Note 19); Grand-Clément, cited above (Note 19).

21 Grand-Clément, cited above (Note 19).

22 For a detailed overview of the issue, see Joaquin Quiñonero-Candela et al., *Dataset Shift in Machine Learning* (Cambridge, MA: The MIT Press, 2022).

23 Maaike Verbruggen, "No, Not That Verification: Challenges Posed by Testing, Evaluation, Validation and Verification of Artificial Intelligence in Weapon Systems," in *Armament, Arms Control and Artificial Intelligence: The Janus-Faced Nature of Machine Learning in the Military Realm*, eds. Thomas Reinhold and Niklas Schörnig (Cham: Springer Nature, 2022), 175–91; Ioana Puscas, AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures (Geneva: UNIDIR, 2023), 22–26, https://unidir.org/wp-content/uploads/2023/10/UNIDIR_AI-international-security_understanding_risks_paving_the_path_for_confidence_building_measures.pdf

24 James Johnson, "The AI Commander Problem: Ethical, Political, and Psychological Dilemmas of Human-Machine Interactions in AI-Enabled Warfare," *Journal of Military Ethics* 21, no. 3–4 (2022): 246–71.

25 Puscas, cited above (Note 23).

26 System cards are documents that report the intended uses and limitations of AI models. They can also provide the results of red teaming exercises. An example is the system card of the Generative Pre-trained Transformer 4 (GPT-4) released by OpenAI. See OpenAI, "GPT-4 System Card," March 23, 2023, https://cdn.openai.com/papers/gpt-4-system-card.pdf

27 Shoker et al., cited above (Note 3); Furkan Gursoy and Ioannis A. Kakadiaris, "System Cards for AI-Based Decision-Making for Public Policy," arXiv:2303.04754 [cs.CY], arXiv, March 1, 2022, https://arxiv.org/abs/2203.04754

28 Track II diplomacy typically involves experts or influential individuals who engage in dialogue on crucial matters but do not represent official capacities.

29 Nadibaidze, cited above (Note 3).

30 Horowitz and Sharre, cited above (Note 3); Favaro, cited above (Note 3); Shoker et al., cited above (Note 3).

31 Holst, cited above (Note 12), 3.

32 Jürgen Altmann, "Confidence and Security Building Measures for Cyber Forces," in *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace*, ed. Christian Reuter (Wiesbaden: Springer Vieweg, 2019), 197.

33 Verbruggen, cited above (Note 23).

34 OSCE, "2019 OSCE Annual Police Experts Meeting: Artificial Intelligence and Law Enforcement—an Ally or Adversary?," https://www.osce.org/event/2019-annual-police-experts-meeting

35 Eliska Pirkova et al., Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual, eds. Deniz Wagner and Julia Haas (Vienna: OSCE Office of the Representative on Freedom of the Media, 2021), https://www.osce.org/files/f/documents/8/f/510332_1.pdf

36 OSCE, "Artificial Intelligence Poses Risks but Can Also Contribute to More Open and Inclusive Societies, Say Participants at ODIHR Event," October 6, 2023, https://www.osce.org/odihr/554413

37 OSCE, "OSCE Court of Conciliation and Arbitration Moot Court Explores Space Activities and Artificial Intelligence," November 16, 2022, https://www.osce.org/court-of-conciliation-and-arbitration/531383; OSCE, "Moot Court in the

Framework of MUNLAWS Conference," https://www.osce.org/court-of-conciliation-and-arbitration/553960

38 See: OSCE, "Panel of Eminent Persons," https://www.osce.org/networks/pep; OSCE, "OSCE Security Days," https://www.osce.org/sg/secdays

39 OSCE Parliamentary Assembly, Luxembourg Declaration (Luxembourg: July 4–8, 2019), 4, https://www.oscepa.org/en/documents/annual-sessions/2019-luxembourg/3882-luxembourg-declaration-eng/file; European Union, OSCE Forum for Security Co-operation N°955, Vienna, 23 September 2020, EU Statement on New Technologies, FSC.DEL/207/20 (Vienna: September 25, 2020), https://www.osce.org/files/f/documents/5/5/466311.pdf; European Union, OSCE Forum for Security Co-operation N°975, Vienna, 12 May 2021, EU Statement on Challenges of New Generation Warfare, FSC.DEL/173/21 (Vienna: May 14, 2021), https://www.osce.org/files/f/documents/7/a/487063.pdf

40 OSCE, "Inter-Regional Conference on the Impact of Emerging Technologies on International Security and Democracy, in the OSCE Asian Partnership for Co-Operation Group Framework," https://www.osce.org/partners-for-cooperation/asian/544219

41 The terms "software" and "hardware" refer to the instructions run by a computer and the physical components constituting the computer system, respectively. In a UAV equipped with AI for autonomous navigation, for example, the software would be the algorithms, while the hardware would be the UAV itself.

42 Altmann, cited above (Note 32).

43 Borghard and Lonergan, cited above (Note 13), 23–24.

44 For a comprehensive list of OSCE cyber confidence-building measures, see OSCE, Decision No. 1202 OSCE Confidence-Building Measures to Reduce the Risks of Conflict Stemming from the Use of Information and Communication Technologies, PC.DEC/1202 (March 10, 2016), https://www.osce.org/pc/227281

45 Verbruggen, cited above (Note 23).

46 Holst, cited above (Note 12).

47 Panel of Eminent Persons on European Security as a Common Project, Back to Diplomacy: Final Report and Recommendations of the Panel (Panel of Eminent Persons on European Security as a Common Project, 2015), 15, https://www.osce.org/files/f/documents/2/5/205846.pdf

48 Nadibaidze and Miotto, cited above (Note 18).

49 Altmann, cited above (Note 32).

50 OSCE, cited above (Note 5), 7.

51 OSCE, cited above (Note 5), 8.

52 OSCE, cited above (Note 5), 9.

53 According to VD11 Chapter XI, "Annual Implementation Assessment Meeting," the AIAM is to be held each year. The VD11 also encourages the holding of "periodic" military seminars, without specifying how often they should take place. See OSCE, cited above (Note 5), 9.

54 OSCE, cited above (Note 5), 13.

55 OSCE, cited above (Note 5), 13.

56 OSCE, cited above (Note 5), 18–19.

57 The author would like to express his gratitude to Lara Maria Guedes and Andrea Miotto for discussions on the implications of artificial intelligence for international security. He is also grateful to Anna Nadibaidze, Argyro Kartsonaki, two anonymous reviewers and the language editor for their valuable feedback on previous drafts. All opinions expressed in this paper are those of the author alone and do not reflect the positions of the OSCE.