sipri

AUTONOMOUS WEAPON SYSTEMS AND AI-ENABLED DECISION SUPPORT SYSTEMS IN MILITARY TARGETING

A Comparison and Recommended Policy Responses

ALEXANDER BLANCHARD AND LAURA BRUUN

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

SIPRI is an independent international institute dedicated to research into conflict, armaments, arms control and disarmament. Established in 1966, SIPRI provides data, analysis and recommendations, based on open sources, to policymakers, researchers, media and the interested public.

The Governing Board is not responsible for the views expressed in the publications of the Institute.

GOVERNING BOARD

Stefan Löfven, Chair (Sweden)
Dr Mohamed Ibn Chambas (Ghana)
Ambassador Chan Heng Chee (Singapore)
Dr Noha El-Mikawy (Egypt)
Jean-Marie Guéhenno (France)
Dr Radha Kumar (India)
Dr Patricia Lewis (Ireland/United Kingdom)
Dr Jessica Tuchman Mathews (United States)

DIRECTOR

Dan Smith (United Kingdom)

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

Signalistgatan 9
SE-169 70 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org

AUTONOMOUS WEAPON SYSTEMS AND AI-ENABLED DECISION SUPPORT SYSTEMS IN MILITARY TARGETING

A Comparison and Recommended Policy Responses

ALEXANDER BLANCHARD AND LAURA BRUUN

June 2025

sipri

STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

© SIPRI 2025
DOI: https://doi.org/10.55163/YQBY3151

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, without the prior permission in writing of SIPRI or as expressly permitted by law.

Contents

Acknowledgements iv
Executive summary v

1. Introduction 1
Scope of this report 2
Outline 2

2. Characterization 3
Functionality of AWS and AI-DSS 3
Impact on the decision to use force 4
Scope of the targeting cycle 5
Role in the execution of targeting decisions 6
Policy implications 6
Figure 2.1. The scope of autonomous weapon systems and AI-enabled decision support systems in the phases of the targeting cycle 4

3. Risks of unintended harm 8
Reliability issues 8
Differences in human–machine interaction 8
Contextual variables 10
Policy implications 11

4. Legal aspects 12
What IHL requires from humans and permits of machines in the conduct of hostilities 12
Legal responsibility and accountability 13
Role in legal assessments 13
Implications for Article 36 reviews 14
Policy implications 14

5. Policy responses 16
Approach 1: Specifically include AI-DSS in existing multilateral efforts on AWS 16
Approach 2: Establish a new process dedicated to AI-DSS 18
Approach 3: Not take a specific approach to AI-DSS 18

6. Key findings and recommendations 20
Key findings 20
Recommendations 20
Figure 6.1. Differences between AI-enabled decision support systems and autonomous weapon systems for human–machine interaction, risk pathways and use of force 21
Table 6.1. Autonomous weapon systems (AWS) and AI-enabled decision support systems (AI-DSS) compared in four key areas 22

About the authors 24

Acknowledgements

SIPRI and the authors express their sincere gratitude to the ministries for foreign affairs of the Netherlands, Sweden and Switzerland for their generous financial support for this publication.

The authors also thank SIPRI colleagues Marta Bo, Vincent Boulanin, Netta Goussac, Dustin Lewis and Jules Palayer for their invaluable insights and support throughout the writing of this report. The authors are also grateful for comments, discussions and feedback provided at different stages of the project by Anna Andersson, Rupert Barrett-Taylor, Nehal Bhuta, Ingvild Bode, Jessica Dorsey, Martin Hagström, Sally Longworth, Arthur Holland Michel and Anna Nadibaidze. Finally, the authors acknowledge the invaluable editorial work of the SIPRI Editorial Department.

Executive summary

The humanitarian and legal concerns raised by autonomous weapon systems (AWS) have long been the subject of international policy processes, and more recently in discussions on the military adoption of artificial intelligence (AI). Growing attention to the military use of AI-enabled decision support systems (AI-DSS) raises the need to consider how these systems fit within global policy conversations. This report compares AWS and AI-DSS for targeting in terms of their respective characterization, risk of unintended harm, legal aspects and policy responses. The report makes a number of key findings to inform policymakers on this issue.

Both AWS and AI-DSS used in military targeting impact the role of humans in targeting decisions. However, a key difference is their scope of use in the targeting cycle: AWS are limited to the mission execution phase, while AI-DSS are used more broadly across multiple phases. Still, the distinction between the two systems can blur in practice, depending on how they are deployed.

Both AWS and AI-DSS carry risks of unintended harm, but these risks emerge in different ways. While both systems share reliability issues arising from known technical limitations of autonomy and AI, their distinct forms of human–machine interaction can lead to different outcomes. For AWS, which have a direct path between target identification and engagement, the risks are direct—for example, a false target identification can result in immediate lethal action without human input. For AI-DSS, which provide outputs to humans, the risk is indirect—harm materializes if humans act upon that false target identification.

Both AWS and AI-DSS raise questions about how much users are permitted to rely on these systems for fulfilling IHL obligations, and how to ensure responsibility and accountability. However, they pose distinct legal challenges. For AWS, concerns stem from the autonomous use of force and whether users can reasonably foresee and control the system's effects. For AI-DSS, concerns arise from humans over-relying on AI when making legal assessments, because of, for example, automation bias, which can potentially lead to situations where the human becomes a passive approver of system recommendations. The use of AI-DSS also raises implications for legal frameworks beyond IHL, such as international human rights law.

These comparisons suggest three approaches available to policymakers navigating the current multilateral context regarding military AI: (*a*) specifically include AI-DSS in multilateral efforts on AWS; (*b*) establish a new process dedicated to AI-DSS; (*c*) take no specific approach to AI-DSS. Each option comes with certain implications and trade-offs that policymakers must take into account. The substantive similarities and differences between AWS and AI-DSS justify any one of these approaches; ultimately, the approach taken will need to reflect political and institutional appetite.

Based on these findings, the report makes three recommendations. First, states should consider whether a dedicated multilateral process for AI-DSS is needed, recognizing the trade-offs between different approaches. Second, future policy efforts on AI-DSS should build on insights from AWS governance, particularly regarding issues of human–machine interaction and oversight, legal compliance and accountability. Third, because AI-DSS also raise distinct issues that are absent or under-explored in AWS processes, policymakers should identify key knowledge gaps and commission research to guide responsible integration of AI into military decision making and support human agency in decisions to use force.

1. Introduction

Autonomous weapon systems (AWS), capable of selecting and applying force to targets without human intervention, alter the human role in the use of force and raise a range of humanitarian, legal and ethical concerns.[1] AWS are, accordingly, the subject of sustained national, regional and international policy debate.[2] The foremost international forum for this debate is the United Nations where, under the auspices of the Meeting of High Contracting Parties to the Convention on Certain Conventional Weapons (CCW) in Geneva, states debate appropriate governance responses to AWS.[3] AWS also feature prominently in a burgeoning global policy conversation on the military adoption of artificial intelligence (AI), including in the First Committee of the UN General Assembly, the Responsible AI in the Military Domain (REAIM) summits and the United States' Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.[4] Although AI is one among many technological enablers of AWS, AWS have long provided a clear, prominent and dramatic example of a use-case of AI in the military domain.

Policymakers are increasingly recognizing that the military adoption of AI is a multifaceted issue, encompassing much more than the issues associated with AWS. This results not least from the interest generated by reported uses of AI-enabled decision support systems (AI-DSS) in contemporary armed conflicts, including by Israel in Gaza and by belligerents in the Russia–Ukraine war.[5] AI-DSS are computerized tools integrating AI that provide information to assist military decision making, including targeting decisions.[6] While these systems are intended to improve military targeting—by augmenting situational awareness, for instance—their capacity to shape targeting decisions, coupled with known limitations of AI, means they also raise humanitarian, legal and ethical concerns.[7]

[1] Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017), pp. 24–27; and International Committee of the Red Cross (ICRC), 'ICRC position on autonomous weapon systems', 12 May 2021.

[2] Blanchard, A. et al., 'Dilemmas in the policy debate on autonomous weapon systems', SIPRI Commentary, 6 Feb. 2025.

[3] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects (CCW Convention), opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983.

[4] See, respectively, UN General Assembly Resolution 79/239, 24 Dec. 2024; REAIM, 'Blueprint for action', 10 Sep. 2024; and US Department of State, Bureau of Arms Control, Deterrence and Stability, 'Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy', 9 Nov. 2023.

[5] AI Now Institute, 'Safety and war: Safety and security assurance of military AI systems', 25 June 2024; Davies, H. and Abraham, Y., 'Revealed: Israeli military creating ChatGPT-like tool using vast collection of Palestinian surveillance data', *The Guardian*, 6 Mar. 2025; Abraham, Y., '"Lavender": The AI machine directing Israel's bombing spree in Gaza.' *+972 Magazine*, 3 Apr. 2024; Israel Defense Forces (IDF), 'The IDF's use of data technologies in intelligence processing', Press release, 18 June 2024; Bendett, S., 'Roles and implications of AI in the Russian–Ukrainian conflict', *Russia Matters*, 20 July 2023; Farnell, R. and Coffey, K., 'AI's new frontier in war planning: How AI agents can revolutionize military decision-making', Belfer Center for Science and International Affairs, 11 Oct. 2024; Hunder, M., 'Ukraine collects vast war data trove to train AI models', Reuters, 20 Dec. 2025; Bondar, K., 'Understanding the military AI ecosystem of Ukraine', Center for Strategic and International Studies (CSIS), 2024; and Bergengruen, V., 'How tech giants turned Ukraine into an AI war lab', *Time Magazine*, 8 Feb. 2024.

[6] Boulanin, V., 'Risks and benefits of AI-enabled military decision-making', eds R. Geiß and H. Lahmann, *Research Handbook on Warfare and Artificial Intelligence* (Edward Elgar: Cheltenham, 2024); Baxter, C. A., 'Enhancing tactical level targeting with artificial intelligence', *Field Artillery*, vol. 1 (2024); Nadibaidze, A., Bode, I. and Zhang, Q., *AI in Military Decision Support Systems: A Review of Developments and Debates* (Center for War Studies: Odense, Nov. 2024); Holland Michel, A., *Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-making* (ICRC: Geneva, 2024); and Probasco, E. et al., 'AI for military decision-making: Harnessing the advantages and avoiding the risks', Center for Security and Emerging Technology (CSET) Issue Brief, Apr. 2025.

[7] ICRC, 'Submission to the United Nations Secretary-General on artificial intelligence in the military domain', 17 Apr. 2025.

This raises a key policy question: how can the global policy debate on military adoption of AI and autonomy, which currently focuses mostly on AWS, address AI-DSS? That is, should AI-DSS and AWS be considered in tandem under the same set of forums and processes, or dealt with separately, in distinct forums or processes? To support policymakers navigating this question, this report compares AWS and AI-DSS in terms of their characteristics, risks of unintended harm and legal aspects. It examines whether—and if so, to what extent—AI-DSS and AWS raise similar or different technical, legal and policy challenges. It aims to make a twofold contribution: first, to inform policymakers in multilateral processes on AWS about the potential suitability of including AI-DSS in such discussions; and second, to inform policymakers involved in international policy initiatives on military AI governance about useful approaches to addressing AI-DSS, including what lessons can be learnt from the global policy debate on AWS.

Scope of this report

AI-DSS can be used for a wide range of military purposes—including supply chain logistics and maintenance, conflict forecasting, and wargaming—but their use for targeting is seen as a high-risk military application of AI.[8] Accordingly, this report focuses on AI-DSS used for 'military targeting', a term broadly understood as 'the use of force by warring parties—whether States' armed forces or organized armed groups—against individuals or objects outside of their control', and encompasses deliberate and dynamic application of force, in offence or defence.[9]

This report does not provide a comprehensive overview of all issues pertaining to AWS and AI-DSS. Rather, it aims to foster a more substantive comparison between the two technologies by limiting the focus to areas and issues subject to significant attention in international policy discussion on AWS, and where there is existing policy expertise, such as humanitarian risks, international humanitarian law (IHL) and human–machine interaction. There are other issues—for example, international human rights law, security issues and ethical issues—not considered in this report but that deserve further attention with respect to both AWS and AI-DSS.

This report is informed by previous SIPRI research, a review of academic and policy literature, and consultations with select experts.

Outline

To set the groundwork for understanding whether AI-DSS and AWS require joint or distinct policy approaches, three chapters undertake a comparative analysis. Chapter 2 begins by defining AWS and AI-DSS, and then compares how their respective technical characteristics impact their role in targeting decisions. Chapter 3 examines how the risks of unintended harm manifest in the use of AWS and AI-DSS in military targeting. Chapter 4 compares legal aspects, including to what extent AWS and AI-DSS pose similar or different challenges, in relation to compliance with IHL. Then, using these comparative analyses, Chapter 5 outlines three policy options to deepen the understanding of whether AI-DSS and AWS require similar or distinct risk mitigation measures. Finally, Chapter 6 presents key findings and recommendations.

[8] Boulanin, 'Risks and benefits of AI-enabled military decision-making' (note 6)

[9] ICRC, 'Targeting under international humanitarian law', *How Does Law Protect in War?*, ICRC online casebook, [n.d.]; and Ducheine, P. A. L., Schmitt, M. N. and Osinga, F. P. B., 'Introduction', eds Ducheine, Schmitt and Osinga, *Targeting: The Challenges of Modern Warfare* (T. M. C. Asser Press: The Hague, 2016).

2. Characterization

This chapter compares the characteristics of AWS and AI-DSS in terms of their functionality and their respective roles in the military targeting cycle. The model of the targeting cycle varies by military doctrine. The model used here (figure 2.1) is a condensed summary of some publicly available doctrines that comprises four broad phases: (1) establishing overarching goals and objectives of the targeting effort; (2) identifying targets and developing target lists, as well as identifying the most appropriate means to attack the target; (3) executing the mission, wherein target persons or objects are identified and attacked; and (4) assessing the effectiveness of the attack.[10] In practice these phases will not always be sequential, but iterative and bidirectional, and sometimes simultaneous.[11] That is, use of force against a target occurs in phase 3, mission execution, but the decision to use force can have origins in phases 1, 2 and 3.

Functionality of AWS and AI-DSS

AWS are weapons that, once activated, can identify, select and apply force to targets without human intervention.[12] AWS function based on pre-programmed target profiles and technical indicators that AWS 'recognize' through their sensors and software. A human operator might supervise and retain the option to override an AWS, but human input is not required for the AWS to function. AWS need not be underpinned by AI to operate (for example, landmines are sometimes classified as AWS). Indeed, the principal multilateral forum on the international governance of AWS, the UN CCW group of governmental experts on emerging technologies in the area of lethal autonomous weapons systems (GGE on LAWS), has mostly taken a technologically neutral stance when discussing these systems—that is, the forum has not taken AI to be an essential aspect of AWS.[13] That said, AWS international policy debate is primarily concerned with advances in AI—such as machine learning (ML) and deep learning (DL)—that would make AWS capable of operating in dynamic, unstructured, open environments and of targeting individual persons.[14]

Decision support systems (DSS) are computerized tools designed to provide information to assist decision making.[15] To do this, they can *indicate* (describe) relevant information, for instance by filtering and flagging information; *predict* (forecast) scenarios, such as enemy troop movement; and *recommend* (prescribe) actions, for example

[10] See e.g. US Air Force, 'Targeting', Airforce Doctrine Publication 3-60, 12 Nov. 2021, p. 7; North Atlantic Treaty Organization (NATO) Standardization Office, 'Allied joint doctrine for joint targeting', NATO Standard AJP-3.9 Edition B Version 1, Nov. 2021, section 1.5; and Australian Defence Force Warfare Centre, 'Targeting', Operations Series ADDP 3.14, 2nd edn, 2 Feb. 2009, sections 1.9 and 4.8–4.9.

[11] See Ekelhof, M. A., 'Lifting the fog of targeting', *Naval War College Review*, vol. 71, no. 3 (2018), p. 66; and Nisser, J., 'Implementing military doctrine: A theoretical model', *Comparative Strategy*, vol. 40, no. 3 (2021).

[12] Boulanin and Verbruggen (note 1), pp. 24–27; ICRC, 'ICRC position on autonomous weapon systems' (note 1); and Taddeo, M., and Blanchard, A., 'A comparative analysis of the definitions of autonomous weapons systems', *Science and Engineering Ethics*, vol. 28 (2022).

[13] See e.g. United Nations, CCW, Group of governmental experts on emerging technologies in the area of lethal autonomous weapons systems (GGE on LAWS), Report of the 2023 session, CCW/GGE.1/2023/CRP.2, 6 May 2023, para. 18.

[14] Sauer, F., 'Military applications of machine learning and autonomous systems', ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume I Euro-Atlantic Perspectives* (SIPRI: Stockholm, May 2019), pp. 86–87.

[15] Bohanec, M., 'Decision support', eds D. Mladenić et al., *Data Mining and Decision Support: Integration and Collaboration* (Springer Science+Business Media: New York, 2003); Şuşnea, E. 'Decision support systems in military actions: Necessity, possibilities and constraints' *Journal of Defense Resources Management*, vol. 3, no. 2 (2012); Holland Michel, *Decisions, Decisions, Decisions* (note 6), p. 13; and Phillips-Wren, G., 'AI tools in decision making support systems: A review', *International Journal on Artificial Intelligence Tools*, vol. 21, no. 2 (2012).

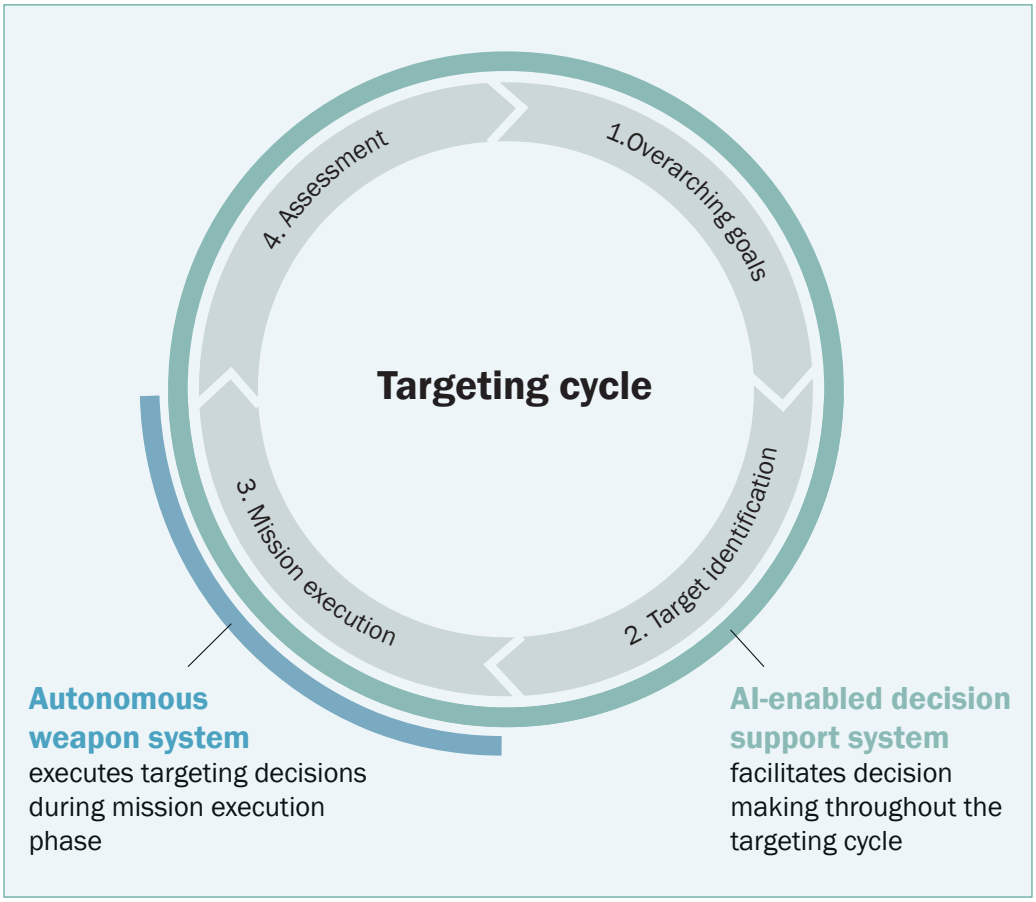

Figure 2.1. The scope of autonomous weapon systems and AI-enabled decision support systems in the phases of the targeting cycle

by presenting an optimal course of action or nominating targets.[16] These are not cumulative functions; DSS can perform one or many of these tasks. DSS can comprise a range of model-based procedures for processing data, including rules-based programming based on formal logic used in so-called expert systems.[17] DSS are not new but current global debate on their military uses is spurred by reported uses of contemporary statistical learning methods in AI, such as ML, DL and large language models (LLMs), which can process and aggregate multiple data sources.[18] AI-DSS are this subset of DSS enabled by contemporary methods in the field of AI. The role of AI-DSS in the targeting context is to mediate some aspects of the users' relationship with the battlespace, by organizing, filtering and presenting information that will inform an assessment about launching an attack.

Impact on the decision to use force

AWS and AI-DSS are both capabilities that impact the human role in the use of force, in part by altering the *process* of making a decision to use force. The decision-making process that leads to the application of force is neither straightforward nor linear, but

[16] ICRC and Geneva Academy, *Expert Consultation Report on AI and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts* (ICRC: Geneva, Mar. 2024), p. 9; and Gadepally, V. N. et al., 'Recommender systems for the department of defense and intelligence community', *Lincoln Laboratory Journal*, vol. 22, no. 1 (2016).

[17] Şuşnea (note 15); and Buchanan, B. G. and Smith, R. G., 'Fundamentals of expert systems', *Annual Review of Computer Science*, vol. 3, no. 1 (1988).

[18] King, A., 'Digital targeting: Artificial intelligence, data, and military intelligence', *Journal of Global Security Studies*, vol. 9, no. 2 (June 2024).

a complex series of interdependent decisions across various practices and processes within defence organizations, involving multiple decision-makers.[19] Introducing machine autonomy into these decision-making processes adds to this complexity and transforms the roles and responsibilities of humans involved, affecting how decisions need to be made. For example, both AWS and AI-DSS can require targeting-related decisions—including generalized target types or models—to be made at lifecycle stages earlier than deployment, including at design and development stages. A distinction, however, is that AI-DSS can shape and influence human decisions about applying force. For example, an AI-DSS used for target prioritization influences what potential targets are seen and what gets focused on by the user.

Scope of the targeting cycle

A difference between AWS and AI-DSS is the scope of the targeting cycle to which each applies. AWS perform a limited number of functions at the mission execution phase (3) of the targeting cycle: identifying, selecting and engaging targets, whereas AI-DSS have a more extensive application across the targeting cycle (figure 2.1).

The use of AI-DSS at phase 1 to meaningfully support higher-order operational intent is likely currently more of an aspiration than an actuality.[20] However, strategic-level uses of AI-DSS—including for wargaming—may indirectly influence the formation of higher-level operational goals in targeting.[21] Phases 2, 3 and 4 are more likely to encompass current, imminent and near-future uses of AI-DSS.

AI-DSS are reportedly already in use at phase 2 for target nomination and prioritization. An example is generating target lists through inferring correlates across data points, such as where an individual can be indicated as a potential target if data on their pattern-of-life behaviours correlates with that of other suspected target individuals. There have also been reported cases of the development and adoption of AI-DSS LLMs to facilitate attack planning.[22] For example, one US tech startup has advertised an AI chatbot, purportedly trained on a dataset that includes military doctrine and IHL, to facilitate the choice of munition for the destruction of a military target.[23] Some commentators have also suggested that AI-DSS can be used at phase 2 to enhance civilian harm mitigation by augmenting collateral damage estimates, including through mapping critical infrastructure.[24]

AI-DSS can also be used at the mission execution phase (3) of the targeting cycle. This might include the use of AI to support selecting and tracking targets, including fixing targets geographically so as to maintain up-to-date awareness on their whereabouts. For example, Israel's 'Where's Daddy?' system reportedly uses mobile phone location data to alert its operators when a pre-determined target is at a given location.[25] AI-DSS can also support mission execution in other ways, including through course-of-action

[19] Ekelhof, M., 'AI is changing the battlefield, but perhaps not how you think: An analysis of the operationalization of targeting law and the increasing use of AI in military operations', eds Geiß and Lahmann (note 6).

[20] Vestner, T., 'From strategy to orders: Preparing and conducting military operations with artificial intelligence', eds Geiß and Lahmann (note 6).

[21] Knack, A., Balakrishnan, N. and Clancy, T., 'Applying AI to strategic warning', Centre for Emerging Technology and Security (CETAS) Research Report, Mar. 2025).

[22] NATO, 'NATO acquires AI-enabled warfighting system', Press release, 14 Apr. 2025.

[23] Scale AI, 'Fine-tuned LLMs for defense', [n.d.]; Biddle, S., 'Meta-powered military chatbot advertised giving "worthless" advice on airstrikes', *The Intercept*, 24 Nov. 2024; and Vincent, B., 'Scale AI unveils "Defense Llama" large language model for national security users', *DefenseScoop*, 4 Nov. 2024.

[24] Greipl, A. R., 'Artificial intelligence in urban warfare: Opportunities to enhance the protection of civilians?', *Military Law and the Law of War Review*, vol. 61, no. 2 (2023); and Tucker, P., 'Special operators hope AI can reduce civilian deaths in combat', Defense One, 26 Aug. 2024).

[25] Abraham (note 5).

analysis or resource optimization, such as indicating the availability or appropriateness of a weapon system.

Finally, AI-DSS can be used to aid post-assessments of an attack at phase 4, including informing battle damage assessments and after-action reports.[26]

Role in the execution of targeting decisions

AWS and AI-DSS not only differ in the scope of their targeting tasks, but also in their role in the execution of a targeting decision. AWS put into effect the identification, selection and engagement of a target without the need of intervention from a human operator. That is, AWS *execute* targeting decisions, based on some form of generalized human decision making made in advance. These generalized decisions can include target types which are then particularized by the AWS during an attack. In contrast, the outputs of AI-DSS are interpreted and acted upon by the user, who makes the particular targeting decision. That is, AI-DSS *facilitate the execution* of targeting decisions by human users.

This difference reflects the distinct operational needs that AWS and AI-DSS are intended to fulfil. AWS are intended to increase operational reach, persistence and speed.[27] For example, AWS capable of operating without a constant communication link can extend operations to environments where there is communication denial or latency, or that are deemed too risky for military personnel.

AI-DSS are intended to support the human decision-maker by 'improving the quality, increasing the speed, and bolstering the consistency of human decisions'.[28] They do this by reducing the need for human analysts to interpret quantities of data that exceed human cognitive capacities for sense-making.[29] Examples include presenting users with a wider range of prescriptive options as part of course-of-action analysis, and filtering and flagging information relevant to a specific target. Both uses relieve the user of the need to manually review intelligence materials and afford the opportunity for exercising higher-order thinking.

Policy implications

AWS and AI-DSS are similar in that they impact decisions about using force, but have two important differences: AWS are used in a narrower set of targeting tasks than AI-DSS, and AWS execute targeting decisions while AI-DSS support humans in making them. In practice, however, AWS and AI-DSS are not always divisible into discrete systems. This is because the functionality of a system can be determined by its context—its configuration, rules of engagement and operational command—rather than by its intrinsic properties. To give an example, the US MIM-104 Patriot missile system has both manual and 'auto-fire' modes. In manual mode the task of engaging the detected and selected target remains with the human operator, while in 'auto-fire' mode the system engages the selected target independently of the human operator.[30]

Systems like the Patriot call into question the utility of a categorical distinction between AWS and AI-DSS, particularly at mission execution stage. Policy deliberation

[26] Hsu, J., 'Battle-damage detector can help aid groups rapidly respond during war', *New Scientist*, 31 May 2024; and Tucker, P., 'How AI could predict the damage to Ukraine from Russian missiles', Defense One, 9 Jan. 2023.

[27] Boulanin and Verbruggen (note 1), pp. 61–63; and ICRC, 'ICRC position on autonomous weapon systems' (note 1).

[28] Holland Michel, *Decisions, Decisions, Decisions* (note 6), p. 18.

[29] Weinbaum, C. and Shanahan, J. N. T., 'Intelligence in a data-driven age', *Joint Force Quarterly*, vol.90, no. 3) (2018), pp. 5–6.

[30] Scharre, P., *Army of None: Autonomous Weapons and the Future of War* (W.W. Norton & Company: London, 2018), chapter 9.

on these systems might need to acknowledge that categorical distinctions between AWS and AI-DSS have to be dynamic or use-case specific. For example, policy debate could focus on how and when human users interact with the systems, including how the distribution of targeting tasks or functions between human and machine changes over the course of a targeting cycle.

More generally, in thinking about the integration of AI into targeting practices, there ought to be attention to (*a*) the conditions under which a system operates as an AWS or an AI-DSS; (*b*) the system's affordances—that is, the hardware or software specifications determining the range of possible actions of the system including, for instance, connectivity to a weapon system; and (*c*) the authorization thresholds and control architectures that shape those roles. Such analysis may need to be more operationally grounded. It is in this context that chapter 3 explores the comparative risks of unintended harm that AWS and AI-DSS pose.

3. Risks of unintended harm

AWS and AI-DSS both carry risks of unintended harm, including harm to civilians. This chapter compares how and where such risks emerge with each system. It focuses on how technical limitations affect both system reliability issues and human–machine interaction, and considers other contextual variables.

The limitations of AI in the military domain are well documented. First, is the 'black box' issue—referring to system opacity, and the (current) dearth of methods to evaluate when and how systems might fail—which presents difficulties for scrutinizing these systems and accounting for unintended behaviours.[31] Second, AI systems require quality (relevant, complete and accurate) data to function reliably, and there are many factors that can affect data quality in the battlefield.[32] AI systems are also susceptible to adversarial behaviours, particularly at the point of deployment, including through data poisoning.[33] Another technical limitation of AI systems is that they are prone to fail when used for tasks or environments different from those for which they were designed.[34] They can, moreover, contain biases, resulting in differential effects along lines of gender and race.[35] AI-DSS, by definition, and AWS that integrate AI will both be liable to these technical limitations.

Reliability issues

At the heart of the risks posed by AWS and AI-DSS is the shared problem of technical reliability.[36] Reliability is the property that the system will function as intended for a given amount of time in a given environment.[37] All technologies, digital or otherwise, present reliability issues that can lead them to fail or to behave in unintended ways. Reliability is particularly pertinent to systems integrating AI because its specific technical limitations, described above, lead it to fail in ways that, for the human user, are highly unpredictable or barely imperceptible, or both—sometimes with faults only noticeable once significant harm has occurred.[38]

Differences in human–machine interaction

Direct or indirect realization of harm

AWS execute targeting decisions independently of a human operator through actuators (i.e. hardware). The 'real-world' materialization of risk can therefore be seen as a direct one. For example, if an AWS identifies a non-threat (e.g. a civilian) as a threat (e.g. a

[31] Bunch, K. et al., *Risk Assessment of Reinforcement Learning AI Systems: Looking Beyond the Technology* (RAND Corporation: Santa Monica, CA, 2024); AI Now Institute (note 5); and Panwar, R. S., Qiang, L. and Shanahan, J. N. T. (eds), 'Military artificial intelligence test and evaluation model practice', INHR and Centre for a New American Security, Dec. 2024.

[32] Holland Michel, A., 'The black box, unlocked: Predictability and understandability in military AI', United Nations Institute for Disarmament Research, 2020, p. 7; and Menthe, L. et al., *Understanding the Limits of Artificial Intelligence for Warfighters: Volume 1—Summary* (RAND Corporation, Santa Monica, CA, 2024).

[33] Hoffman, W. and Kim, H. M., 'Reducing the risks of artificial intelligence for military decision advantage', CSET Policy Brief, Mar. 2023; and Uesato, J. et al., 'Adversarial risk and the dangers of evaluating against weak attacks', *Proceedings of Machine Learning Research*, vol. 80 (2018).

[34] Taddeo, M. et al., 'Artificial intelligence for national security: The predictability problem', CETAS Research Report, Sep. 2022).

[35] Blanchard, A. and Bruun, L., 'Bias in military artificial intelligence', SIPRI Background Paper, Dec. 2024.

[36] Holland Michel, *Decisions, Decisions, Decisions* (note 6), p. 18; and Probasco et al. (note 6).

[37] International Organization for Standardization, 'Trustworthiness—Vocabulary', ISO/IEC TS 5723:2022, July 2022.

[38] Ryan, M., 'In AI we trust: Ethics, artificial intelligence, and reliability', *Science and Engineering Ethics*, vol. 26, no. 5) (2020).

combatant)—known as a false positive—the system will directly target the non-threat, unless something prevents the attack.

Risk mitigation measures for AWS have thus tended to focus on managing the challenge of unintended behaviours.[39] These measures fall into two categories. The first focuses on exercising control over the AWS before the deployment or mission execution phase, by ensuring system design includes control parameters such as target types, duration of operation and geographical range of deployment.[40] These measures also include auditability of AWS, systems of certification, and dynamic updating of rules of engagement.[41] The second set of measures cover forms of supervisory control for operators to intervene, such as real-time kill switches.[42]

In contrast, AI-DSS are intended to support human decision making, so the manifestation of 'real-world' risk is, in principle, indirect. That is, a human must both interpret and then act upon information (outputs) provided by the AI-DSS, for risk to materialize as harm. For example, if an AI-DSS present a human with inaccurate, incomplete, misleading or false information—for example, misrepresenting a non-threat as a threat—harm only materializes if the human both accepts that information to be true and acts upon it. AI-DSS risk mitigation therefore must focus on whether users can challenge, correct or disregard inaccurate outputs.

Effects on human judgement

Since AWS execute targeting decisions at a distance—geographic, temporal, epistemic—from their human operators, users must assess the likely effect of the system in the context of use. This assessment is key to informing judgements about the legality and ethicality of a given use of the system. Such assessments will always include an irreducible amount of uncertainty. Unintended harm can result from the use of an AWS because the user could not foresee, or was not required to foresee, all possible eventualities resulting from the system's deployment.

Because AI-DSS mediate some aspects of their users' relationship with the battlespace, they present a different risk scenario.[43] AI-DSS organize, filter and present information that can inform assessments about launching an attack. AI-DSS will be detrimental to the decision-making process if they distort these assessments by presenting inaccurate, incomplete, misleading or false information. For example, an AI-DSS that provides inaccurate information about the presence of civilians or civilian objects will potentially lead to the system user(s) making an inaccurate assessment regarding the risk of harm to civilians or civilian objects in an attack. This risk is exacerbated by automation bias, where 'humans place greater confidence and trust in the outputs of automated systems than their own critical deliberative skills'.[44] It is important to note that these risks may result from design choice, rather than human error or tech-

[39] Blanchard, A. et al., 'A risk-based regulatory approach to autonomous weapon systems', *Digital Society*, vol. 4, no. 23 (2025).

[40] Boulanin, V. et al., 'Limits on autonomy in weapon systems: Identifying practical elements of human control', SIPRI/ICRC, June 2020.

[41] Cummings, M. L., 'Lethal autonomous weapons: Meaningful human control or meaningful human certification?', *IEEE Technology and Society Magazine*, vol. 38, no. 4 (2019); Kwik, J. et al., 'Controlling military artificial intelligence: Harnessing rules of engagement and military directives', T. M. C. Asser Institute Policy Brief No. 2025-01, 3 Feb. 2025; and Spayne, P. et al., 'Operating itself safely: Merging the concepts of "safe to operate" and "operate safely" for lethal autonomous weapons systems containing artificial intelligence', *Defence Studies*, vol. 25, no. 1 (2025).

[42] Verdiesen, I., Santoni de Sio, F. and Dignum, V., 'Accountability and control over autonomous weapon systems: A framework for comprehensive human oversight', *Minds & Machines*, vol. 31 (2021); and Sharkey, N., 'Towards a principle for the human supervisory control of robot weapons', *Politica & Società*, vol. 3, no. 2 (2014).

[43] Boulanin, 'Risks and benefits of AI-enabled military decision-making' (note 6).

[44] Bode, I., 'Human–machine interaction and human agency in the military domain', Centre for International Governance Innovation (CIGI), Policy Brief no. 193, Jan. 2025, p. 6.

nical limitation. AI-DSS are, after all, intended to provide a reductive representation of the battlespace—for example, by filtering complexity and noise—to highlight features relevant to facilitating targeting practices.[45]

AI-DSS can also contribute to the materialization of unintended harm by increasing the speed at which humans make targeting decisions. Speed is a key strategic driver for military adoption of AI to support decision making, but speed is also a key source of risk.[46] Increased speed means more attacks can happen in quick succession or before more or better information about the target is obtained. That is, speed can impinge on the capacity for humans to exercise oversight of AI-DSS outputs. This provides greater opportunity for harm to materialize, for instance by exposing civilians to a higher number of attacks in a short period or before their presence is discovered.

The twin concerns of speed and automation bias introduce the risk that rather than 'supporting' the human decision-maker, AI-DSS come to exert an outsized and undue influence on the decision-making process. At best, this may mean that AI-DSS become detrimental to the decision-making process, resulting in sub-optimal outcomes. At worst, it means that AI-DSS are used as de facto AWS, where the hierarchy between human operator and AI-DSS is flipped, with the operator exercising only nominal oversight over outputs and passively approving system recommendations. The effect is that the space for humans to exercise judgement over AI-DSS may be diminished or closed entirely.

Contextual variables

In practice, the materialization of unintended harm in the use of AWS and AI-DSS will depend not just on the intrinsic properties of the systems, but also contextual variables that affect how the systems will perform and will be used. Three sets of factors are relevant here, all of which have quantitative and qualitative effects on the risks of unintended harm: operating conditions, system affordances and control architectures. These factors either multiply the routes through which risk can emerge, or minimize the friction on the route by which risk can materialize, principally by diminishing the scope for humans to exercise (meaningful) oversight.

Operating conditions

Operating conditions are the environmental properties under which the system is used. Variables that affect system performance and user ability to foresee risk include the extent to which the environment is known in advance, observable, cluttered, and dynamic or static. For example, the use of either AWS or AI-DSS in environments (relatively) absent of civilians or civilian objects reduces risk of harm to civilians or civilian objects.

System affordances

System affordances include, but are not limited to, a system's software, hardware and connectivity to other systems—for example, whether an AI-DSS is used as a standalone system or is connected to a weapons platform. In some experimental programmes, for instance, AI-DSS are connected to real-time fire-control systems so that the AI can cue

[45] Barrett-Taylor, R., 'The limits of digital representations of the battlefield', CETAS Expert Analysis, June 2024.

[46] Schwarz, E., 'The (im)possibility of responsible military AI governance', ICRC Humanitarian Law & Policy Blog, 12 Dec. 2024; Bo, M. and Dorsey, J., 'Symposium on military AI and the law of armed conflict: The "need" for speed—the cost of unregulated AI decision-support systems to civilians', *OpinioJuris*, 4 Apr. 2024; and Reynolds, I. J., 'Speed and war in US military thought: Mapping the conditions for AI-enabled decision-making', *Millennium* (1 Apr. 2025).

a weapons platform automatically for the human user.[47] This direct connectivity to weapons hardware reduces the friction between the AI-DSS output and the materialization of risk.

Control architectures

Control architectures shape the use of AWS and AI-DSS in the targeting cycle. They include the organizational and technical configurations that determine who has the capacity to authorize the use of these systems, under what conditions, and with what level of automation.[48] Control architectures are an essential part of ensuring oversight of automated systems in targeting. For example, the Habsora ('Gospel') system used by Israel in Gaza operations reportedly embeds AI to recommend targets based on aggregated intelligence inputs. Formal strike authorizations remain with the human operator, but the system is reportedly geared towards rapid target recommendation cycles in a way that minimizes friction between target recommendation and target engagement. Without a well-calibrated control architecture, a system like this could tip towards the functional automation of targeting decisions, even if control remains formally with a human.

Control architectures can include the processes and procedures that enable users to test, evaluate and guarantee the performance of a system.[49] They can also include those that establish parameters of use and risk mitigation measures that respond to system limitations, such as data governance practices and organizational workflows.[50] For AI-DSS, organizational workflows might include verifying outputs at relevant stages of the targeting cycle, and assigning responsibilities for weighting and appropriately communicating outputs in broader intelligence assessments.[51]

Policy implications

While reliability issues for AWS and AI-DSS overlap, the way these systems interact with human decision making differ, leading to the risks of unintended harm emerging in different ways. Understanding these differences helps assess where risk mitigation strategies might align or diverge. Three key contextual elements—environmental conditions, system affordances and control architectures—also shape how risk materializes for both systems. Future policy efforts, especially regarding AI-DSS, could focus on these elements to better manage and reduce the risk of unintended harms.

[47] Northrop Grumman, 'Northrop Grumman adds cutting-edge AI capabilities to forward area air defense', Press release, 7 Oct. 2024; Zoldi, D., 'Countering rogue UASs with modular AI- and ML-enabled systems', Military Embedded Systems, 2 Mar. 2022; Insinna, V., 'US Air Force awards contracts for drone wingman's AI brains, but keeps details secret', *Breaking Defense*, 29 July 2024; and Airbus, 'Airbus and Helsing to collaborate on artificial intelligence for the teaming of manned and unmanned military aircraft', Press release, 5 June 2024.

[48] Devitt, S. K., 'Meaningful human command: Advance control directives as a method to enable moral and legal responsibility for autonomous weapons systems', eds Geiß and Lahmann (note 6); Boulanin, V. and Lewis, D. A., 'Responsible reliance concerning development and use of AI in the military domain', *Ethics and Information Technology*, vol. 25, no. 1 (2023); and Bo, M., Bruun, L. and Boulanin, V., *Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS* (SIPRI: Stockholm, Oct. 2022).

[49] US Department of Defense, Chief Digital and Artificial Intelligence Office (CDAO), *Test and Evaluation of Artificial Intelligence Models: What to Consider in a Test & Evaluation Strategy* (CDAO: Washington, DC, Apr. 2024).

[50] Afina, Y. and Grand-Clément, S., 'Bytes and battles: Inclusion of data governance in responsible military AI', CIGI Paper no. 308, Oct. 2024; and Vogel, K. M. et al., 'The impact of AI on intelligence analysis: Tackling issues of collaboration, algorithmic transparency, accountability, and management', *Intelligence and National Security*, vol. 36, no. 6 (2021).

[51] Hughes, M. et al., 'AI and strategic decision-making: Communicating trust and uncertainty in AI-enriched intelligence', CETAS Research Report, Apr. 2024.

4. Legal aspects

States agree that the development and use of both AWS and AI-DSS must comply with international law, including international humanitarian law (IHL).[52] In the policy debate on AWS, compliance with IHL has been central, with states developing a nuanced understanding of the challenges AWS pose for IHL.[53] This chapter explores whether the legal issues raised by AI-DSS overlap with those of AWS, and to what extent insights from the policy debate on AWS can inform the policy efforts on AI-DSS.

What IHL requires from humans and permits of machines in the conduct of hostilities

AWS and AI-DSS used in targeting both engage the set of IHL rules that regulate the conduct of hostilities, including the principles of distinction, proportionality and precautions in attack.[54] Also, both technologies provoke questions about what these rules fundamentally require of humans and permit from machines.

This interpretive question has been at the core of the AWS debate. It is broadly recognized that IHL does not prohibit humans from relying on automated technologies to support, inform and implement aspects of the evaluative decisions and judgements required to comply with the rules regulating the conduct of hostilities.[55] But it is also widely accepted that the *extent* to which IHL permits such reliance is not unlimited, because human agency must be retained in targeting decisions. While the exact terms are debated, 'human agency' is here used to reflect the broad view that, to satisfy their legal obligations and ensure accountability, humans must retain the ability to reasonably *foresee* and *control* the behaviour and effects of any weapon system.[56] Since AI-DSS also introduce questions about the extent to which humans are permitted to rely on machine processes while performing conduct of hostilities obligations, lessons learnt from the AWS policy debate on these issues can be applied to the AI-DSS context. That is, assumptions about the importance of retaining human agency established in the AWS debate can be used as a baseline approach to identify (im)permissible types and degrees of human–machine interaction in the context of AI-DSS. The International Committee of the Red Cross (ICRC) and others have already indicated the importance of ensuring human agency in the context of AI applications in the military domain, including AI-DSS.[57]

[52] CCW, GGE on LAWS, Report of the 2019 session, CCW/GGE.1/2019/3, 25 Sep. 2019, annex IV, 'Guiding principles'; UN General Assembly Resolution 79/239 (note 4); and REAIM (note 4).

[53] See e.g. CCW, GGE on LAWS, CCW/GGE.1/2023/CRP.2 (note 13).

[54] Protocol Additional to the 1949 Geneva Conventions, and relating to the Protection of Victims of International Armed Conflicts (AP I), opened for signature 12 Dec. 1977, entered into force 7 Dec. 1978, Arts 41, 48, 51, 52 and 57; and ICRC, 'Rules', Customary IHL Database, [n.d.], Rules 1, 6, 7, 13,14, 15, 16, 17, 18, 19 and 47.

[55] See e.g. CCW, GGE on LAWS, 'Implementing IHL in the use of autonomy in weapon systems', Working paper submitted by the United States, CCW/GGE.1/2019/WP.5, 28 Mar. 2019; CCW, GGE on LAWS, 'United Kingdom proposal for a GGE document on the application of IHL to emerging technologies in the area of lethal autonomous weapon systems (LAWS)', Proposal by the United Kingdom, Mar. 2022; ICRC, 'Submission to the United Nations Secretary-General on artificial intelligence in the military domain' (note 7); and Bruun, L., Bo, M. and Goussac, N, *Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What Does IHL Permit, Prohibit and Require?* (SIPRI: Stockholm, Mar. 2023).

[56] See e.g. CCW, GGE on LAWS, CCW/GGE.1/2023/CRP.2 (note 13); ICRC, 'ICRC position on autonomous weapon systems' (note 1); Bruun, L., *Towards a Two-tiered Approach to Regulation of Autonomous Weapon Systems: Identifying Pathways and Possible Elements* (SIPRI: Stockholm, Aug. 2024), p. 19; and Lewis, D. and Sweeny, H., 'Exercising cognitive agency: A legal framework concerning natural and artificial intelligence in armed conflict', Harvard Law School Program on International Law and Armed Conflict, Legal Concept Paper, Jan. 2025.

[57] ICRC, 'Submission to the United Nations Secretary-General on artificial intelligence in the military domain' (note 7); Peace Movement Aotearoa and Stop Killer Robots (Aotearoa New Zealand), 'Submission to UN Secretary-

Legal responsibility and accountability

Responsibility and accountability for making legal determinations rest with humans (state agents as well as individuals) and cannot be transferred to a machine.[58] That applies to both AWS and AI-DSS. However, concerns have long been raised that AWS may undermine states' abilities to investigate and repress harmful incidents (as required by IHL), consequently undermining the ability to ensure accountability.[59] These concerns flow especially from two factors. First is the 'black box' factor, which refers to the challenges of understanding and explaining on what basis complex automated systems produce outputs.[60] Second, is the 'many hands problem', where the behaviours of an AWS likely reflect decisions made by a variety of actors at multiple stages of the system's lifecycle, from its design and development to its use.[61]

Similar issues around accountability can also arise in the context of AI-DSS. AI-DSS are likely to introduce challenges around technical and organizational traceability, making it potentially difficult to understand and explain on what basis specific outputs—analyses, recommendations and predictions—came about.[62] Because AI-DSS have a broad scope of application, these issues of traceability could manifest in multiple ways and at various (earlier) points across the targeting cycle. For example, reliance on AI-generated outputs in early phases of the targeting process may lead to increased opacity around how decisions to apply force came about. However, how AI-DSS impact the ability to hold humans accountable is an area that deserves further attention by experts and policymakers.

Role in legal assessments

With AWS, users make legal assessments, including those regarding distinction and proportionality, *before* activating the system. The concern, therefore, is whether such assessments in relation to a specific attack will remain valid throughout the attack—that is, whether users will be able to reasonably foresee and control the behaviour and effects of AWS upon activation to ensure compliance with IHL.[63] One particular concern is that the use of AWS in dynamic environments or over long periods could render assessments made pre-activation obsolete.

In contrast, AI-DSS do not directly engage the same concerns around the ability to reasonably foresee and control the behaviour and effects of these systems. Unlike AWS, AI-DSS do not 'execute' legal assessments but are instead designed to *inform* them, and their behaviours are in principle always produced under human supervision. However, this does not mean that AI-DSS are unproblematic when it comes to compliance with

General on AI in the military domain', 11 Apr. 2025; and Bode I. et al., 'Submission to the UN Secretary-General on AI DSS in the military domain', 9 Apr. 2025.

[58] CCW, GGE on LAWS, 'Guiding principles' (note 52); UN General Assembly Resolution 79/239 (note 4); and REAIM (note 4).

[59] Protocol AP I (note 54), Arts 85 and 146. On AWS and accountability see Bo, M., 'Autonomous weapons and the responsibility gap in light of the mens rea of the war crime of attacking civilians in the ICC statute', *Journal of International Criminal Justice*, vol. 2, no. 19 (2021); Crootof, R., 'War torts', *New York University Law Review*, vol. 97, no. 4 (2022); and Geiß, R., 'State control over the use of AWS: Risk management and state responsibility', eds R. Bartels et al., *Military Operations and the Notion of Control under International Law* (T. M. C. Asser Press: The Hague, 2021).

[60] See e.g. ICRC, *ICRC Position on Autonomous Weapon Systems and Background Paper* (ICRC: Geneva, May 2021), p. 7; and CCW, GGE on LAWS, 'State of Palestine's proposal for the normative and operational framework on autonomous weapon systems', Working paper submitted by Palestine, CCW/GGE.1/2023/WP.2, 3 Mar. 2023.

[61] Bo, Bruun, and Boulanin (note 48).

[62] Bo (note 59); and Holland Michel, *Decisions, Decisions, Decisions* (note 6), p. 57.

[63] ICRC, *ICRC Position on Autonomous Weapon Systems and Background Paper* (note 60), p. 7; and Woodcock, T. K., 'Human/machine(-learning) interactions, human agency and the international humanitarian law proportionality standard', *Global Society*, vol. 38, no. 1 (2023).

conduct of hostilities rules. Instead, AI-DSS raise questions about the extent to which humans can rely on AI outputs when making legal assessments.[64] For example, does reliance on probabilistic targeting advice erode the context-specific, precautionary nature of many IHL obligations pertaining to attack? This is an aspect that has not featured strongly in the AWS debate but which must be addressed when discussing how to ensure compliance with IHL in the context of AI-DSS. The ICRC and subject-experts have warned that the use of AI-DSS may undermine users' ability to make the context-specific and value-based human judgements needed to comply with IHL.[65] As discussed in chapter 3, one concern is that the increased speed, distance and biases associated with the use of AI-DSS may result in operators becoming passive approvers of the system's recommendations, rather than exercising their own judgement.[66]

Implications for Article 36 reviews

The fact that AI-DSS 'inform' rather than 'execute' legal assessments can also have implications for the obligation to conduct legal reviews. Under Article 36 of Additional Protocol I of the Geneva Conventions, parties to a conflict are obliged to conduct legal reviews of all new weapons, means and methods of warfare to ensure that their employment in some, or all circumstances, is lawful.[67] As weapon systems, AWS are clearly subject to Article 36 reviews. However, it is less clear to what extent AI-DSS, especially sub-components such as critical software, are subject to such a review.[68] The ICRC has argued that AI-DSS integrated into weapon platforms and those that influence how weapon systems are used should be subject to a legal review as they amount to 'means' and 'methods' of warfare, respectively.[69] However, this issue remains to be systematically addressed by states and experts.

Policy implications

Efforts to ensure compliance with IHL in the context of AI-DSS can, to some extent, draw on the AWS debate: both AWS and AI-DSS are subject to the same rules regulating the conduct of hostilities and their use generates questions around what compliance with these rules requires from humans and permits of machines in targeting decisions. And extensive discussions around ensuring accountability in the context of AWS can, to some extent, be applied to the AI-DSS debate. For example, commonly agreed measures around ensuring responsible chains of command and control, audit trails and mechanisms to investigate harmful incidents are relevant to both AWS and AI-DSS.[70]

[64] ICRC and Geneva Academy (note 16), p. 10; Woodcock (note 63); and Dorsey, J. and Bo, M., 'AI-enabled decision-support systems in the joint targeting cycle: Legal challenges, risks, and the human(e) dimension', *International Law Studies* (forthcoming).

[65] See e.g. ICRC and Geneva Academy (note 16), p. 17; Dorsey and Bo (note 64); Holland Michel, *Decisions, Decisions, Decisions* (note 6); Nadibaidze, Bode, and Zhang (note 6); Scharre (note 30); Boutin, B. et al., 'DILEMA statement on the global governance of artificial intelligence in the military', Designing International Law and Ethics into Military Artificial Intelligence (DILEMA) Project, Asser Institute, Jan. 2023; and Stewart, R. and Hinds, G., 'Algorithms of war: The use of artificial intelligence in decision making in armed conflict', ICRC Humanitarian Law & Policy Blog, 24 Oct. 2023.

[66] Dorsey and Bo (note 64); and Zhou, W. and Greipl, A. R., 'Artificial intelligence in military decision-making: Supporting humans, not replacing them', ICRC Humanitarian Law & Policy Blog, 29 Aug. 2024.

[67] AP I (note 54), Art. 36.

[68] See e.g. CCW, GGE on LAWS, 'Guiding principles' (note 52); and Klonowska, K., 'Article 36: Review of AI decision-support systems and other emerging technologies of warfare', eds Gill, T. D. et al., *Yearbook of International Humanitarian Law*, vol. 23 (The Hague: T. M. C. Asser Press, 2020).

[69] ICRC, 'Submission to the United Nations Secretary-General on artificial intelligence in the military domain' (note 7).

[70] See e.g. CCW, GGE on LAWS, 'Rolling text', 8 Nov. 2024.

However, despite these overlapping issues, the fact that AWS *execute* targeting decisions while AI-DSS *inform* them means the use of AI-DSS introduces distinct challenges, particularly around the extent to which humans are permitted to rely on AI outputs to inform legal assessments.

Moreover, insofar as AI-DSS can be used for a range of tasks beyond force application, discussions about lawful use should be seen as an opportunity to explore legal implications beyond IHL rules on the conduct of hostilities, such as the Geneva Convention rules and obligations related to prisoners of war and the protection of civilians, as well as law other than IHL.[71] For example, the increased surveillance and data collection associated with the use of AI-DSS stress the need to better understand how these practices impact compliance with international human rights law. While other bodies of international law are also relevant to the development and use of AWS, they have been largely underexplored relative to the attention paid to IHL. However, the broader range of military applications for AI-DSS and its growing prominence in various policy forums increases the urgency of exploring the wider implications under international law of increased reliance on AI in the military domain.

[71] Geneva Convention (III) Relative to the Treatment of Prisoners of War, opened for signature 12 Aug., entered into force 21 Oct. 1950; and Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War, opened for signature 12 Aug. 1949, entered into force 21 Oct. 1950.

5. Policy responses

AWS have been the subject of sustained national, regional, and international policy debate for over a decade.[72] The foremost international forum for this debate is the UN CCW GGE on LAWS, established in 2016, where states debate appropriate regulatory responses to AWS. Since 2023, there have been two additional tracks for multilateral discussions on AWS: the First Committee of the UN General Assembly and the REAIM initiative. There is, in short, a robust global multilateral debate on AWS.

In contrast, AI-DSS used in military targeting are a relatively novel topic in global policy conversation, lacking the familiarity among policymakers that AWS enjoy. There is, for instance, currently no dedicated multilateral forum for discussing a regulatory response to AI-DSS. Nevertheless, a number of policymakers and civil society groups have urged for the development of a multilateral response to AI-DSS in light of the risks they pose.

This contrast raises a key policy question: how should global policy debate address AI-DSS? That is, do AI-DSS and AWS require similar or different policy responses, and should these systems be considered separately or in tandem?

The aim of this chapter is to support policymakers facing this question. It outlines three approaches available to policymakers: (*a*) specifically include AI-DSS in existing multilateral efforts on AWS; (*b*) establish a new process dedicated to AI-DSS; or (*c*) take no specific approach to AI-DSS. The chapter draws on the similarities and differences between AWS and AI-DSS highlighted in the previous chapters to assess the feasibility, including possible pitfalls, of each option.

Approach 1: Specifically include AI-DSS in existing multilateral efforts on AWS

One approach available to policymakers is to include AI-DSS in existing multilateral tracks dedicated to AWS, such as the GGE on LAWS. The substantial overlap in issues associated with AI-DSS and AWS provide some justification for this approach. For example, AI-DSS used in military targeting raise issues states have placed at the forefront of multilateral debate on AWS, including meaningful human involvement in the use of force, and risks to civilians and civilian objects.[73] The international community's longstanding focus on AWS can be seen as an opportunity for responding to AI-DSS. A decade of policy debate has provided a substantial pool of knowledge about the automation of military targeting, including the challenges of human–machine interaction with complex automated systems. Indeed, a longstanding insight from AWS policy discussions—that integrating technologies into the targeting process introduces distinct challenges, even apart from the actual use of force—provides a valuable foundation for engaging with the complexities of AI-DSS. Moreover, many of the tools, technologies, and processes associated with AWS may also be relevant to certain forms of AI-DSS.

Including AI-DSS in processes dedicated to AWS could therefore offer significant advantages by leveraging existing costs and resources; and allowing policymakers to benefit more immediately from the expertise accumulated in multilateral efforts on AWS.

There are, however, two main limitations of this approach. First, the technological neutrality of current multilateral efforts on AWS could mean that the specific issues raised by the use of AI are overlooked. Second, the difference in the range of appli-

[72] Blanchard, A. et al., 'Dilemmas in the policy debate on autonomous weapon systems' (note 2).

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

cations of AWS and AI-DSS could undermine the efforts by confining or diluting the topics under discussion. These limitations are detailed below.

Technological neutrality

Existing multilateral forums on AWS, particularly the GGE on LAWS, have for the most part remained neutral about the types of technologies that can underpin AWS. This reflects the fact that current initiatives arose in the context of increased use of drones for targeted killings, and concerns that advances in a range of technologies—not just AI—would enable targeted killings without human oversight.[74] While AI and data issues feature increasingly in the GGE on LAWS, its mandate continues to restrict its scope to 'emerging technologies in the area of lethal autonomous weapon systems'.[75] That is, the focus is on systems that identify, select and engage targets without human intervention.

Current concerns about AI-DSS in military targeting arose as part of wider concerns about the profound societal impact of AI and related data collection and processing technologies. That is, the recent attention given AI-DSS primarily as a consequence of their use in contemporary armed conflict dovetails with a wider global policy debate on safe and responsible uses of AI.

So long as multilateral forums dedicated to AWS continue to take a technologically neutral stance, addressing AI-DSS in these forums would mean cleaving AI-DSS from the types of societal concerns that have placed it on the global policy agenda. That is, failing to focus on the specific technology involved—AI— might mean losing the opportunity to benefit from insights and advances related to the governance of AI in other safety-critical domains such as healthcare and law enforcement. For example, if AI-DSS come within civil AI governance trajectories, there will need to be work to reconcile IHL-focused governance with human rights-based constraints on data use, not least in dual-use surveillance contexts.[76] This could also mean that international policy efforts on AI-DSS become separated from efforts on other potential applications of military AI.

Scope of application

Existing multilateral initiatives on AWS have largely been motivated by the humanitarian concerns raised by the automated application of force. As described above, the focus of the GGE on LAWS has been on systems used in military targeting: AWS. As noted in chapter 2, AI-DSS encompass a much wider range of applications at the strategic, operational and tactical levels of military operations, including for battle management, supply chain logistics, conflict and instability forecasting, and humanitarian services. Even within the scope of military targeting—the focus of this report—AI-DSS cover a much broader range of functions and tasks than AWS.

Including AI-DSS in discussions dedicated to AWS would mean either addressing a very narrow subset of AI-DSS—those systems most closely aligned to AWS—or expanding the scope of discussions in AWS-dedicated forums. The first course of action would result in a very restricted multilateral response to AI-DSS, the second in discussions that either become too superficial to be meaningful, or that require considerable time and resources to cover the range in any detail. The latter path would likely offset any efficiencies gained from leveraging the processes and resources of existing processes.

[74] For example, the report that galvanized contemporary multilateral initiatives on AWS seldom mentions AI, focusing instead on advances in robotics. United Nations, General Assembly, Human Rights Council, 23rd session, 'Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions', A/HRC/23/47, 9 Apr. 2013.

[75] See e.g. CCW, GGE on LAWS, 'Rolling text' (note 70).

[76] See e.g. Morley, J., et al., 'Governing data and artificial intelligence for health care: Developing an international understanding', *JMIR Formative Research*, vol. 6, no. 1 (2022).

Approach 2: Establish a new process dedicated to AI-DSS

A second approach is to establish a dedicated forum in which to address AI-DSS. This could be a dedicated focus on AI-DSS as part of existing processes dedicated to military AI, such as the inclusion of military AI on the agenda of the First Committee of the General Assembly, or in the REAIM initiative. Or, it could be a specific process dedicated to AI-DSS within existing forums, for example by setting up an expert working group or similar under the auspices of the GGE on LAWS.

The substantial differences in issues associated with AI-DSS and AWS— the vast difference in the range of applications of these systems, the different forms of human–machine interaction each presupposes, and the different sets of legal issues each provokes (see chapters 2–4)—provide justification for either pathway under this approach. A dedicated process on AI-DSS would be an opportunity to address these issues and explore a wider range of matters than AWS-dedicated processes allow. Policymakers could integrate insights and advances on the governance of AI from other domains, to address much broader applications of AI-DSS, and to move beyond framings that have impeded discussion—and thus progress—in the GGE on LAWS. There are, however, at least two challenges of this approach: first, the additional resources and expertise needed; and second, the risk of overlapping yet contradictory governance regimes.

Effort, time and resources involved in establishing the parameters of debate and developing expertise

In any new process, it will take time to reach consensus on the parameters of debate. Given the range of AI-DSS types and uses, decisions will need to be made on whether to take a broad view or to focus on high-risk applications. While dedicated governance efforts on AI-DSS should at minimum draw on policy insights and the expertise generated by efforts on AWS, they will need to be supported by additional research on how to translate and apply that expertise to AI-DSS, and by new research into aspects relevant to AI-DSS that are absent or under-explored in discussions on AWS. These might include, for instance, the use of data collection and processing technologies associated with AI, and the human rights considerations these engage.

Risk of overlapping yet contradictory governance regimes

As suggested in chapter 2, the distinction between AWS and AI-DSS is more porous than the categorical labels suggest. At the point of mission execution, at least, the distinction between the two systems is less a reflection of their intrinsic properties than of their conditions of use, system affordances and control architectures. Making the same categorical distinction in the policy debate thus reflects a policy choice, such as the GGE on LAWS choosing to take a technology-neutral approach to AWS. Because AI-DSS and AWS can overlap in practice, separate governance processes dedicated to each type of system could also contain areas of overlap. The risk is that this overlap leads to duplication of effort or inconsistent governance regimes for each system. Separate processes dedicated to AWS and AI-DSS should aim to avoid both outcomes by exchanging expertise and insights on progress, to ensure symmetry and alignment in potentially duplicated or contradictory policy responses.

Approach 3: Not take a specific approach to AI-DSS

A third option is that states do not take a specific approach to AI-DSS. This might mean addressing AI-DSS as part of a larger framing, including within the REAIM initiative, avoiding a system-centric approach. States could therefore omit explicit reference to AI-DSS in ongoing processes, such as the General Assembly's resolution on military

AI, the GGE on LAWS and the REAIM initiatives, with the expectation that their outcomes will nonetheless impact on the development and use of AI-DSS. An upside of this approach is flexibility and the scope to discuss issues related to a wide range of applications of AI in the military domain. A downside is that, without explicit reference to AI-DSS, the relevance of those outputs to AI-DSS are likely to be fairly high-level and lacking detailed measures to mitigate the specific risks posed by AI-DSS.

6. Key findings and recommendations

The results of the foregoing comparative analyses of AWS and AI-DSS in the areas of characterization (chapter 2), risks of unintended harm (chapter 3), legal aspects (chapter 4) and policy responses (chapter 4) are summarized in table 6.1. It shows that while there are some similarities between the two kinds of systems—noting that any categorical distinction is a matter of policy choice, as in practice the systems can overlap in some contexts of deployment—there are more differences, with different implications for the question of whether to address them in tandem or separately. This chapter presents two key findings relating to the comparisons, and three key recommendations for policymakers to advance on this topic.

Key findings

AWS and AI-DSS both impact human involvement in the use of force

AWS and AI-DSS are both capabilities that alter the human role in the use of force, in part by altering *how* decisions about using force are made. The decision-making process that leads to the application of force is already complex and distributed across multiple phases and actors. Introducing AWS and AI-DSS into the various stages of this process raises questions about the roles, responsibilities and legal obligations of human decision-makers across this process. For example, both AWS and AI-DSS raise concerns about how human oversight and control over these systems is maintained and how to minimize the risk of unintended harm. For AWS, these concerns relate primarily to the system operating at a distance without direct user input; for AI-DSS, to the increased speed with which targeting decisions can be made and the extent to which users rely on the system's outputs without challenging or correcting outputs.

AWS and AI-DSS fulfil distinct functions in the use of force

Understanding the relatively distinct ways in which AWS and AI-DSS impact the human role in targeting decisions is important for managing risks of unintended harm and ensuring lawful use (figure 6.1). AWS 'execute' targeting decisions, so the concerns revolve around whether the autonomous application of force will lead to unforeseen outcomes that are legally non-compliant. In contrast, AI-DSS 'inform' targeting decisions, so the concern is that they mediate a user's relationship with the battlespace, by filtering, selecting and analysing data that is presented to the user to interpret and act upon. The way AI-DSS alter the human role in targeting decisions is, in a sense, more complex and wide-ranging than for AWS because AI-DSS *influence* judgements about targeting decisions. This is particularly concerning if AI-DSS present inaccurate, incomplete, misleading or false information about the battlespace.

Recommendations

Consider a multilateral process dedicated to AI-DSS

Given high-profile reported uses of AI-DSS in contemporary armed conflicts, some policymakers and civil society groups have called for urgent development of a multilateral response to AI-DSS. There are at least three options available to states in responding to these calls: (*a*) include AI-DSS in multilateral processes dedicated to AWS; (*b*) establish a new process dedicated to AI-DSS, either separately or within AWS processes; or (*c*) take no specific approach to AI-DSS, in the hope that current processes

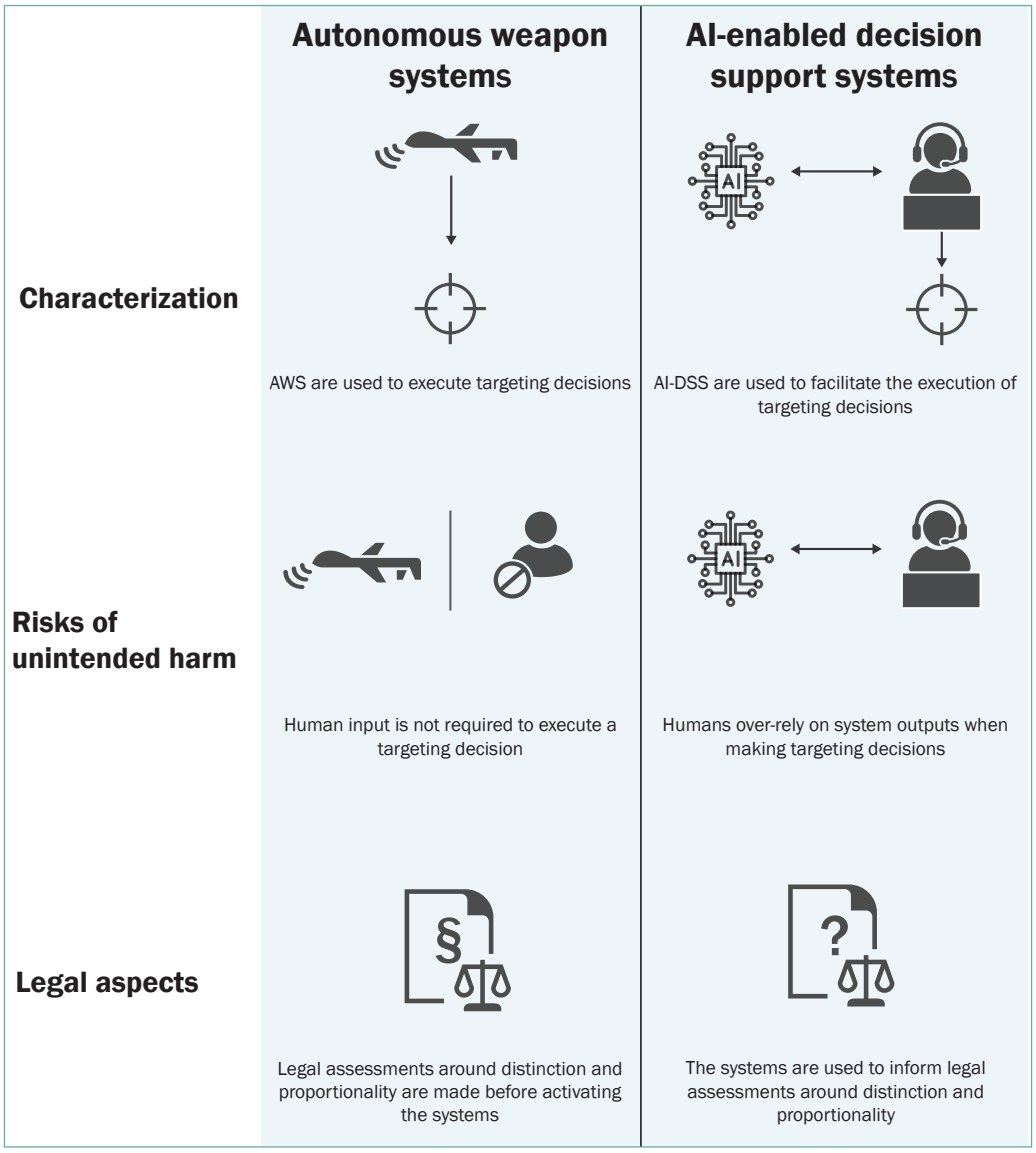

Figure 6.1. Differences between AI-enabled decision support systems and autonomous weapon systems for human–machine interaction, risk pathways and use of force

will address AI-DSS. Each of these options brings trade-offs and feasibility issues. States ought to begin deliberating on how best to respond to these calls.

Policy efforts and expertise on AWS need to inform policy and complement expertise on AI-DSS

The international community's longstanding focus on AWS can be seen as an opportunity for addressing AI-DSS. Policy efforts on AWS have provided a substantial pool of knowledge about human–machine interaction in targeting, including how to ensure that humans and institutions fulfil their legal obligations, foresee and mitigate risk, and can be held accountable. In light of these overlapping themes, international governance efforts on AI-DSS should draw on insights generated in the AWS context. Policymakers ought to begin exploring about how these insights can translate to and complement efforts on AI-DSS.

Research is required to increase understanding about AI-DSS

AI-DSS raise a number of issues and considerations that have not been given detailed treatment, or have been absent, from international policy efforts on AWS. This includes,

Table 6.1. Autonomous weapon systems (AWS) and artificial intelligence-enabled decision support systems (AI-DSS) compared in four key areas

Area	Similarities	Differences
Characterization	Both AWS and AI-DSS transform human involvement in targeting decisions.	1. Scope of the targeting cycle: AWS are limited to the mission execution phase, while AI-DSS are used more broadly across multiple phases. 2. Role in the execution of the targeting decision: AWS execute targeting decisions, while AI-DSS facilitate the execution of targeting decisions. 3. AI-DSS have a wider range of applications beyond military targeting.
Risks of unintended harm	Both AWS and AI-DSS have reliability issues arising from known technical limitations of AI.	1. Differences in human–machine interaction: With AWS, risks are direct because human input is not required to execute a targeting decision, whereas with AI-DSS, the risk is indirect because a human must interpret and act on its outputs. 2. Effects on human judgement: AWS users must assess the likely effect of the system in the context of use, so risk arises from an inability or failure to foresee harm, whereas AI-DSS users must interpret, accept and act on information it presents, so risk arises from poor data, flawed outputs and automation bias.
Legal aspects	1. Both AWS and AI-DSS are subject to international humanitarian law (IHL) rules in the conduct of hostilities. 2. Both raise questions around user reliance on machines for fulfilling IHL obligations, and how to ensure responsibility and accountability.	1. Role in legal assessments: Because AWS 'execute' legal assessments, concerns stem from whether users can reasonably foresee and control the system's effects in advance. AI-DSS present concerns about to what extent humans may rely on AI outputs when making legal assessments. 2. Article 36 reviews: As weapons technology, AWS are subject to legal reviews, but it is less clear whether AI-DSS are subject to the same requirement. 3. Beyond IHL: Like many AI systems, AI-DSS have implications for human rights law (e.g. surveillance and data collection) and environment law (e.g. sustainability of component extraction and production), whereas these aspects, where they apply to AWS, are under-explored.
Policy responses	1. Both AWS and AI-DSS require policy responses that address the role of machine autonomy in the use of force. 2. To the extent that AWS incorporate AI and AI-DSS are used for military targeting, both require policy responses that address the use of AI in military targeting.	1. Technological neutrality: Multilateral efforts on AWS arise from concerns about weapons that can identify, select and engage targets without human intervention, keeping the focus technologically neutral. Concerns about AI-DSS arise in the context of wider concerns about AI as a technology, beyond military applications. 2. Scope of application: AWS are used only in military targeting, but AI-DSS have broader applications, both civil and military, that need a wider set of considerations.

for instance, discussions on the use of data collection and processing technologies associated with AI, and the human rights considerations these systems engage. The fact that AI-DSS is a nascent issue on the global policy agenda also provides a prime opportunity for states to consider *how* the integration of AI into consequential military decision-making processes is framed. To this end, AI-DSS could present an opportunity to discuss and develop policy responses about how to support or enable the exercise of human agency in targeting decisions that involve systems integrating AI. States ought to identify and commission research on keys areas that need addressing.

About the authors

Dr Alexander Blanchard is a Senior Researcher in the SIPRI Governance of Artificial Intelligence (AI) Programme. His work focuses on issues related to the development, use and control of military applications of AI.

Laura Bruun is a Researcher in the SIPRI Governance of AI Programme. Her focus is on how emerging military technologies, notably autonomous weapon systems and military AI, affect compliance with—and interpretation of—international humanitarian law (IHL). Laura's work focuses, among others, on how bias in military AI systems impacts compliance with IHL and how to ensure human responsibility in the development and use of military AI.

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE
Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org

© SIPRI 2025