EXPERT CONSULTATION REPORT

ARTIFICIAL INTELLIGENCE AND RELATED TECHNOLOGIES IN MILITARY DECISION-MAKING ON THE USE OF FORCE IN ARMED CONFLICTS

CURRENT DEVELOPMENTS AND POTENTIAL IMPLICATIONS

REPORT

GENEVA ACADEMY
Académie de droit international humanitaire et de droits humains
Academy of International Humanitarian Law and Human Rights

ICRC

EXPERT CONSULTATION REPORT

ARTIFICIAL INTELLIGENCE AND RELATED TECHNOLOGIES IN MILITARY DECISION-MAKING ON THE USE OF FORCE IN ARMED CONFLICTS

CURRENT DEVELOPMENTS AND POTENTIAL IMPLICATIONS

Report drafted by Anna Greipl, Research Assistant, Geneva Academy of International Humanitarian Law and Human Rights, with contributions from Neil Davison and Georgia Hinds, International Committee of the Red Cross (ICRC).

Citation: ICRC and Geneva Academy, *Expert Consultation Report on AI and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts*, Geneva, ICRC, March 2024.

CONTENTS

EXECUTIVE SUMMARY

During the past decade, there has been a significant focus on autonomous weapon systems (AWS) in legal and policy discussions on the application of artificial intelligence (AI) in the military domain. In recent years, however, awareness has been increasing of the fact that the military applications of AI are much broader.[1]

Against the backdrop of these discussions and debates about the opportunities and risks of using AI for military purposes, the ICRC and the Swiss Chair of International Humanitarian Law (IHL) at the Geneva Academy of International Humanitarian Law and Human Rights held two online expert workshops in November 2022, with the aim of **increasing our understanding of the role of military applications of AI-based decision support systems (AI DSS) in decision-making on the use of force in armed conflicts – including, but not limited to, the specific decision tasks of target selection and the application of force**.

The workshops focused in particular on decisions regarding the application of force in armed conflict, due to their significant impact on peoples' lives and dignity, and on communities. Such decisions also give rise to specific obligations under IHL.

As armed forces explore the use of increasingly complex AI techniques, with the aim of using DSS to support ever more complicated military decision-making tasks related to the use of force, they also need to consider how to preserve the human judgement necessary to ensure compliance with IHL and uphold protection for civilians and combatants alike.

This report highlights and examines some of the themes that arose during the above-mentioned workshops. These findings do not necessarily represent the views of the Geneva Academy, the ICRC or individual expert participants. Addressed primarily to political decision makers, academics, researchers and lawyers, the report aims to provide a preliminary understanding of the challenges and risks related to the use of AI DSS in military decision-making on the use of force, and to explore what measures may need to be implemented, with regard to the design and use of AI DSS, to mitigate risks to those affected by armed conflict.

The main takeaways from the expert consultations are:

The introduction of AI to DSS for military decision-making on the use of force in armed conflicts adds a new dimension to existing challenges relating to non-AI-based DSS. The use of these systems raises new questions regarding the understandability and predictability of their outputs, the speed at which they operate, and barriers to the human ability to effectively assess the accuracy of their outputs. As a result, the use of AI DSS capabilities has the potential to reduce the human judgement involved in military decision-making on the use of force in armed conflicts, thus raising humanitarian, legal and ethical questions.

The military application of AI DSS and of AWS raises distinct legal and conceptual issues, even though some of the underlying technology may be very similar. For instance, some applications of AI DSS, such as those developed for automated target recognition, could form part of an AWS. However, once a weapon system carries out the process of selecting and applying force to a target without human intervention (i.e. an AWS), distinct humanitarian, legal and ethical concerns arise. This has prompted the ICRC to recommend a specific regulatory response, in the form of new, legally binding rules on AWS.[2]

1 International Committee of the Red Cross, "ICRC Position Paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach" *International Review of the Red Cross,* March 2021.

2 International Committee of the Red Cross, "ICRC position on autonomous weapon systems", 12 May 2021 available online at https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems.

The realities of warfare mean that the challenges and risks linked to the use of AI DSS in other contexts will likely be exacerbated when AI DSS are used in military decision-making on the use of force in armed conflicts. This raises particular concerns, given the significant potential impact of these kinds of decisions on peoples' lives and dignity, and on communities.

Preserving human judgement in military decision-making on the use of force in armed conflicts is crucial to reducing humanitarian risks, addressing ethical concerns and facilitating compliance with IHL. This may require new approaches to existing challenges arising from the interaction between humans and AI DSS, as well as technical requirements relating to the use of AI DSS. Some approaches may help to address existing technical challenges (e.g. predictability, understandability, and bias), while others may help to improve human decision makers' ability to critically engage with and use AI DSS outputs (e.g. mitigating automation bias). Nevertheless, to ensure the viability of such approaches, it is important to assess them in relation to their use in the specific context of armed conflict, and to require that their use is restricted to tasks, and contexts, for which they have been specifically and rigorously tested.

The use of AI DSS in military decision-making on the use of force may require additional measures and constraints to reduce risks for people affected by armed conflicts, and to facilitate compliance with IHL. Many of the existing challenges of human-machine and human-AI interaction are likely to persist, while certain technical limitations may be insurmountable. Therefore, it may be necessary to place certain constraints on the use of AI DSS in decisions relating to the use of force. These may include: restricting the use of AI DSS to certain tasks or decisions and/or to certain contexts; placing specific constraints on the use of AI DSS with continuous learning functions, due to their more unpredictable nature; and slowing down the military decision-making process at certain points to allow humans to undertake the qualitative assessments required under IHL in the context of specific attacks.

THE WAY FORWARD

The expert consultations conducted as part of this project reveal the need to pursue additional research and dialogue in this area, in order to better understand the measures and constraints that may be required with regard to design and use of AI DSS, to mitigate the risk of harm to people affected by armed conflict, and to ensure compliance with IHL. Further analysis will be needed to identify the applications of AI DSS in this context that have the biggest impact on decisions on the use force.

INTRODUCTION

This report is part of a joint initiative, entitled "Digitalization of Conflict: Humanitarian Impact and Legal Protection", by the ICRC and the Swiss Chair of International Humanitarian Law at the Geneva Academy of International Humanitarian Law and Human Rights. As part of this initiative, two expert online workshops were held in November 2022, with the aim of **assessing the application of AI decision support systems (AI DSS) in military decision-making on the use of force, from a humanitarian perspective.** A range of experts from different parts of the world, with relevant professional backgrounds in different fields, participated in their individual capacity, alongside representatives from the ICRC and the Geneva Academy.[3]

This report highlights and expands on some of the themes that arose during these consultations (for further information on the issues raised, please see: Holland Michel, Arthur, *Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making*, ICRC, March 2024). The Geneva Academy and the ICRC are the sole authors of the report, whose findings do not necessarily represent the consensus, or individual views, of the experts consulted. Addressed primarily to political decision makers, academics, researchers and lawyers, the report aims to provide a preliminary understanding of the challenges and risks related to the use of AI DSS in military decision-making on the use of force, and to explore possible ways to address them.

To this end, the report is divided into five interrelated sections. Section 1 develops the conceptual framework. Section 2 explores the key drivers of the military application of AI DSS in decision-making on the use of force, and offers examples of current developments and uses of such applications. Section 3 explores some of the risks posed by the use of AI DSS in military decision-making on the use of force, in terms of compliance with IHL and the potential impact on people affected by armed conflict. Section 4 contains an initial assessment of ways to mitigate these risks, with respect to the technical aspects of the system and the human-machine interaction process. Section 5 summarizes the main findings and puts forward possible next steps. A background document prepared for the workshops is provided in an annex to this report.

3 For a list of experts consulted during the workshops, see Annex 2.

SECTION 1

UNDERSTANDING ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS IN MILITARY DECISION-MAKING ON THE USE OF FORCE

Discussions on the military applications of AI have begun to pay increasing attention to the use of AI DSS in decision-making on the use of force, beyond the tasks of target selection and the application of force.[4] This section aims to clarify the meaning of *AI DSS in military decision-making on the use of force* for the purposes of this report.

WHAT DO WE MEAN BY MILITARY DECISION-MAKING ON THE USE OF FORCE?

Military decision-making processes that lead to the use of force are complex. These critical decisions involve multiple actors at various command levels, from the higher political level down to the strategic, operational and tactical levels.[5]

At the *political level*, the political leadership sets the objectives and guidance that will apply to all other levels of decision-making. Then, at the *strategic level*, political aims are translated into more specific military objectives and guidelines, including important decisions related to the use of force. For instance, the rules of engagement set out the circumstances, conditions, degree and manner in which force may be applied. Furthermore, at this level, the types of targets that may be engaged are set out, as well as the number of civilian casualties that a military operation may sustain without seeking approval at the highest levels. At the *operational command* level these objectives and guidelines are then turned into specific tasks for the tactical forces. This process includes multiple critical decisions on the use of force, including the analysis, selection and prioritization of targets, as well as weaponeering and collateral damage assessments. At the *tactical level*, planners assess target attributes, the use of weapons and their respective impacts, and any potential unforeseen outcomes. They also decide on whether the deployment of equipment and military personnel needs to be modified to comply with legal, doctrinal or other requirements and take into account conditions on the ground. In carrying out a military assignment, information and intelligence are collected and evaluated to identify and track targets, or to proceed with collateral damage assessments to inform decisions on whether to attack, or to suspend or cancel an attack.

Most military decisions may be considered as directly or indirectly relevant to the use of force. For instance, logistical decisions, such as planning the deployment of personnel or the transport of weapons, equipment, and personnel, are essential to ensuring the effectiveness of military

4 Recent high-level discussions reflect this development. For instance, the REAIM 2023 Summit that took place on 15 and 16 February 2023 aimed to provide a "a platform for all stakeholders (governments, industry, civil society, academia and think tanks) to forge a common understanding of the opportunities, dilemmas and vulnerabilities associated with military AI." For more information, see "About REAIM 2023", (Government of the Netherlands), available online at https://www.government.nl/ministries/ministry-of-foreign-affairs/activiteiten/reaim/about-reaim-2023.

5 While the terminology used here, to an extent, derives from Western military doctrines, most major military forces agree that there are essentially three main levels of command.

operations, and indirectly influence the manner in which force will be applied. Nevertheless, for the purposes of this consultation and report, military decisions on the use of force are defined as those closely related to the time, place and object of an attack. Many of these critical decisions are made well before the final stage of the military operation requiring the use of force, starting with decisions taken at the political level. This is particularly true in deliberate targeting operations that may begin weeks, months or years before the attack on a target takes place, with sufficient time available to apply a strategic approach. While largely guided by the same decision-making steps, in dynamic targeting operations the decision to attack a target is compressed in time so that military forces can act in a more responsive and timely manner to constantly evolving situations.[6]

Overall, it is important to keep in mind that military decision-making processes on the use of force are complex and can start well before a weapon system is deployed or force is used. The following section conceptualizes DSS and, most importantly, the role AI plays in the use of these systems for military decision-making on the use of force.

WHAT ARE ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS FOR MILITARY DECISION-MAKING ON THE USE OF FORCE?

DSS can, broadly speaking, be characterized as computerized tools that are designed to assist humans at different levels in the chain of command to complete decision-making tasks.[7] The main purpose of such systems is to inform military decision makers, and enable the user to make decisions that support effective, rapid and legitimate military reactions or actions.

Traditional, non-AI-based DSS have significant limitations when used in decision support tasks relating to fast-changing or uncertain situations, such as armed conflicts, where relevant variables are difficult to encode. In an attempt to overcome these limitations, a newer generation of DSS now make use of AI to benefit from powerful computing tools to better collect, integrate, manage and analyse large and complex data sets.[8] This report focuses on data-driven AI techniques, such as machine learning and deep learning, which essentially consist of building an AI system by letting it "learn"[9] through experience, in the form of training data. Thus, in contrast to earlier AI DSS, where knowledge was provided by humans in the form of specifically programmed rules or instructions, data-driven learning algorithms can generate their own rules for solving a particular problem or carrying out a task.

A conceptual distinction is often made in legal discussions between DSS designed to assist decision makers and those that are connected to physical elements enabling them to transform output into physical action. A typical example of the latter is an AWS, which (by definition) selects and applies force to targets without human intervention.[10] From a legal perspective, this is an important distinction. However, as we shall see in the following section, even the use of AI DSS presents challenges regarding human-AI interaction, which may give rise to concerns similar to those raised by the use of AWS.

6 Merel Ekelhof and Giacomo Persi Paoli, "The human element in decisions about the use of force", UNIDIR, March 2020 available online at https://unidir.org/publication/the-human-element-in-decisions-about-the-use-of-force/.

7 See background document prepared for the expert workshop in Annex to this report.

8 Anna Rosalie Greipl, "Data-Driven Learning Systems and the Commission of International Crimes: Concerns for Criminal Responsibility?", *Journal of International Criminal Justice*, 2023, p. 5.

9 The term "learning" in this context must be understood in a functional sense: a computer program is capable of learning when it improves its performance in certain tasks through experience in the form of data. See: Stuart J. Russell and others, *Artificial Intelligence: A Modern Approach*, Third edition, Global edition, Pearson, 2016.

10 ICRC *supra* note 1.

Currently, AI DSS are being developed by companies and tested by armed forces to support decision makers by collecting and analysing evidence, detecting familiar patterns in the data, checking hypotheses, suggesting possible courses of action and evaluating the appropriateness of proposed actions. Some of these decision support tasks are descriptive (the system collects, organizes and presents past data to human decision makers), while other tasks are predictive (the system identifies patterns and trends in past data and uses probability to predict possible future outcomes and their likelihood) or even prescriptive (the system recommends the best possible course of action).

One of the major concerns about increasing the development and potential use of data-driven AI DSS for military decision-making on the use of force is that these systems are used to highlight certain data points and to prioritize and select features based on predetermined features and self-learned patterns *without* human intervention. This decrease in human involvement is of particular concern considering the complex characteristics of existing AI DSS, the challenges of human-AI interaction, and the growing pressure on military decision makers to increase the speed of decision-making processes on the use of force in contemporary armed conflicts. It raises humanitarian, legal and ethical questions. Legally, humans are responsible for complying with the law – including making judgements and conducting the assessments required under the IHL rules on the conduct of hostilities – and the resulting accountability cannot be transferred to an AI DSS. Moreover, many of these legal assessments require qualitative, context-specific value judgements. Moreover, ethically speaking, upholding human agency in critical decisions leading to the use of force is necessary to uphold the principles humanity, human dignity and moral responsibility.[11]

Accordingly, the use of increasingly complex AI techniques enabling DSS to execute ever more complicated tasks to support military decision-making on the use of force raises a key humanitarian, legal and ethical question: *how to preserve the necessary human judgement in the use of complex AI DSS outputs?*

CLARIFYING THE SCOPE OF THE DISCUSSION: DISTINGUISHING ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS FROM AUTONOMOUS WEAPON SYSTEMS

The rise of increasingly complex AI techniques in military applications is also a prominent topic in discussions on AWS. While some of the concerns about the use of AI DSS in military decision-making on the use of force and AWS may overlap, there are – from a conceptual and legal perspective – at least two important differences between the two systems.

Firstly, AI DSS may be used at various stages of the military decision-making process, up to and including the decision to use force, rather than only at the final decision point where an AWS would apply force. Secondly, AI DSS are intended to assist or inform – but not replace – decision makers; the decisions to select and apply force to targets are, in practice, still made by humans. By contrast, in the case of an AWS, human intervention is – by definition – removed from the process of selecting and applying force to a target, which is mediated by sensors and/or software.

Consequently, these applications of AI give rise to very different concerns. In the case of AWS, the main concern arises from the fact that the human does not choose a specific target or the precise time or place of attack, which is instead self-initiated or triggered by the system itself. The challenge with AI DSS, on the other hand, involves ensuring that the system helps and supports, rather than hinders or displaces, human decision-making on the use of force, in line with

11 ICRC, *supra* note 1.

the decision makers' obligations under IHL. Accordingly, some of the issues that have arisen during discussions on AWS, such as the "accountability gap" or the extent to which IHL assessments can be carried out by an AI system process, do not arise to the same extent with AI DSS, since humans retain their decision-making role.

Even though – conceptually and legally – these systems raise different concerns, the underlying technologies are potentially very similar. AI DSS systems – specifically those developed for automatic target recognition – could be used to directly trigger the application of force by an AWS. Particular concerns have been raised with regard to AWS based on data-driven learning AI technologies, especially with regard to how they function and the fact that their impact is not sufficiently predictable or understandable.[12]

12 The notion of understandability is used in the following text, encompassing the more technical terms used in the literature on AI, such as explainability or interpretability, while accounting for the human ability to understand AI systems.

SECTION 2

ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS APPLICATIONS IN MILITARY DECISION-MAKING ON THE USE OF FORCE

The desire to develop technical systems capable of assisting humans in their decision-making is nothing new. In this context, the findings of political scientists in the 1950s, such as Herbert Simon, were influential; Simon argued that people may be rational but are limited in their cognitive processing abilities when they have to deal with complex problems.[13] The desire to overcome these human characteristics led to the progressive development of DSS in various fields, including the military domain.

In the light of advances in the development of DSS, supported by new and more powerful AI techniques, various technologically advanced military powers started to pursue and utilize this military capability, including to support decisions on the use of force. Some of the main factors that explain the military interest in integrating AI DSS into decision-making processes on the use of force will be explored in the following section.

WHAT DRIVES THE MILITARY INTEREST IN ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS FOR DECISION-MAKING ON THE USE OF FORCE?

Two of the key drivers of the use of AI DSS are the desire to enhance and accelerate the military decision-making cycle, as compared with present command and control structures. For the armed forces, these provide decisive military advantage in armed conflicts.

One way to conceptualize successful military decision-making is through the observe-orient-decide-act (OODA) loop model (Figure 1) of command and control.[14]

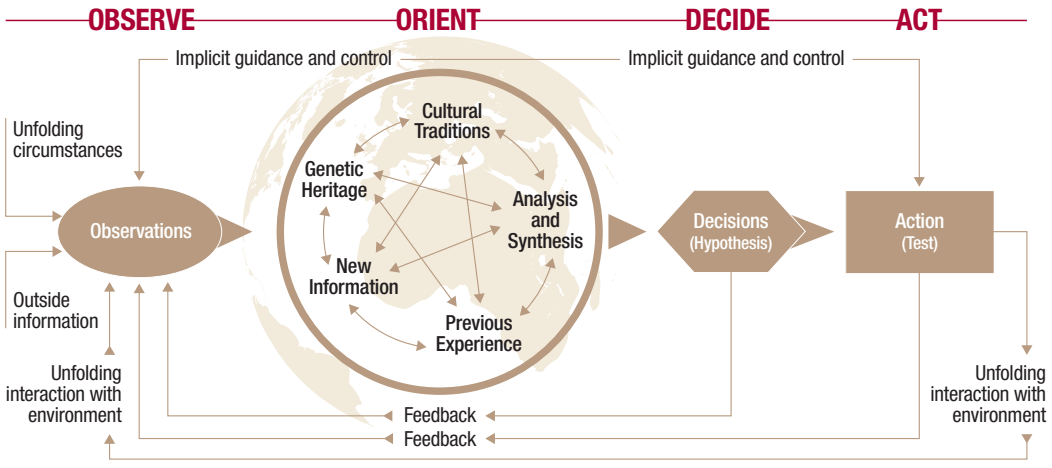

Figure 1: John Boyd's OODA loop diagram

13 Herbert Simon, *Administrative Behavior, 4th Edition*, 1997.

14 John Boyd, *A Discourse on Winning and Losing*, Edited and Compiled by Dr Grant T Hammond, Air University Press, 2018.

Broadly speaking, the "observe" phase involves collecting data on the assignment at hand. During the "orient" phase, the collected information is processed, in order to understand the strategic environment. In the "decide" phase, commanders determine what course of action will produce the desired outcome. Lastly, in the "act" phase, military personnel test the chosen hypothesis by interacting with the environment and then decide whether to execute the selected course of action, or whether more information is needed before any action can be taken. The overall aim of the OODA loop is to execute the decision-making cycle faster than the adversary, to act/react more quickly on the basis of better information, and to outpace and thereby disrupt the OODA loops of the adversary.[15] In essence, the faster the OODA loop, the greater the advantage a military force will have over its opponents.

However, commanders still face a major challenge to accelerating this decision-making cycle: the need to reduce uncertainty, with very limited time to do so. In fact, uncertainty is a fundamental attribute of the dynamic decision-making environment in military operations. Therefore, command and control aims to reduce the number of uncertainties that commanders must deal with, so that they are able to make sound decisions. Theoretically, uncertainty can often be reduced by obtaining additional information on situations. This process requires time, a precious commodity in armed conflicts. In the time it takes to find more information about a particular situation ("observation"), the adversary may already be acting and changing their course of action. In other words, in armed conflicts knowledge rapidly becomes obsolete. Moreover, the rapid tempo of modern operations limits the amount of information that can be gathered, processed and assimilated in time to be of use. Thus, the fundamental challenge for military decision-making arises from the tension between coping with uncertainties and racing against time.[16]

To address the uncertainties of contemporary warfare, armed forces have invested in developing new data collection tools, such as sensor feeds from uncrewed platforms, to enable them to collect massive amounts of raw data from across the globe, 24 hours a day, seven days a week.[17] The volume of data collected using these tools has grown exponentially since the early 2010s, requiring increasing processing capabilities. Data analysts are no longer capable of manually scrutinizing all the raw data generated from reports, sensors, videos and other data feeds. The military decision-making process has thus become an issue of scale: new data collection tools have significantly improved the ability to collect information ("observation") which has – in turn – increased interest in technical solutions to automate the processing of collected data ("orientation").[18]

In the light of this, and since the speed and complexity of conflicts may increase, given the growing urbanization of armed conflicts, as well as large-scale military operations, AI DSS have gained traction among armed forces as a means to help them process large amounts of data and obtain real-time information or recommendations from different domains – including space, air, land, sea and cyber – about their adversary, so that they can act or react rapidly. Despite the technical limitations of existing AI DSS, current developments seem to be driven by the assumption that these limitations can be overcome through technological development.

Along similar lines, military forces often cite the potential for AI DSS to support faster and more accurate decisions on the use of force that better protect civilians and their livelihoods from harm during military operations. This is particularly relevant in urban warfare, where it is considera-

15 In Boyd's view, the goal of military tactics should be to operate in a manner to get inside of the adversary's decision-making cycle (or OODA loop), to "[…] enmesh the adversary in a world of uncertainty, doubt, mistrust, confusion, disorder, fear, panic, chaos, […] and/or fold adversary back inside himself so that he cannot cope with events/efforts as they unfold".

16 Taylor Anderson, *MCDP 6 Command and Control*, CreateSpace Publishing, 2017 available online at https://www.barnesandnoble.com/w/mcdp-6-command-and-control-taylor-anderson/1134752866.

17 Richard H Shultz and Richard D Clarke, "Big Data at War: Special Operations Forces, Project Maven, and Twenty-First-Century Warfare" *Modern War Institute*, 25 August 2020 available online at https://mwi.usma.edu/big-data-at-war-special-operations-forces-project-maven-and-twenty-first-century-warfare/.

18 *Ibid*.

bly more difficult to conduct military operations in a manner that effectively protects civilians, due to the presence of civilians and the interconnected nature of military objectives and civilian objects, including critical infrastructure. In such situations, the capacity of AI DSS to process large data sets has been put forward as an example of how these systems may help commanders to increase their awareness of the presence of civilians and civilian objects.[19]

Finally, armed forces are searching for new tools to ensure that their tactics are effective, in the light of the changing nature of contemporary armed conflict. As a result, AI DSS are often described as key to supporting the armed forces in developing new approaches to combat.

WHAT TYPES OF ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS ARE CURRENTLY BEING DEVELOPED AND/OR DEPLOYED?

This section offers a non-exhaustive list of concrete examples of the types of AI DSS being developed and/or deployed. While these examples include applications at each command level, the expert discussions suggested that the most common near-future applications are likely to be at the operational and tactical levels.

When it comes to the **strategic level**, examples of the development and/or the use of AI DSS relating to the use of force remain relatively limited. Systems may assist decision makers in monitoring the battlefield and developing scenarios. This may include the development of AI DSS to predict the behaviour and reactions of other states or simulate the progress of ongoing conflicts, including war gaming models.

At the operational level, various AI DSS are currently being developed and/or used in decision-making on the use of force, such as the identification, selection and prioritization of potential targets. These assessments ordinarily rely on an intelligence cycle, which consists of processing the raw data collected into intelligence to use to make decisions on the selection of potential targets. To this end, AI DSS may process information by detecting objects and events, as well as by collecting data and generating intelligence.

In the field of object recognition, tremendous progress has been observed in recent years with regard to the use of machine learning techniques in the private and military sectors.[20] Current military research projects are focusing on overcoming the inability of existing systems to infer abstract meanings from visual data by **developing techniques to infer the intentions of individuals or groups, including threats posed by people**, from surveillance imagery.[21] The outputs of these AI DSS may play a critical role in decision-making on the use of force, as they may inform military decision makers' assessment of whether a particular person or object may be considered a lawful target.

19 Note that some scholars and practitioners have questioned the effectiveness of faster and more accurate targeting decisions as a key measure to enhance the protection of civilians. See Damian Copeland and Lauren Sanders, "Engaging with the Industry: Integrating IHL into New Technologies in Urban Warfare", Humanitarian Law & Policy Blog (7 October 2021), available online at https://blogs.icrc.org/law-and-policy/2021/10/07/industry-ihl-new-technologies/; Ruben Stewart and Georgia Hinds, "Algorithms of war: The use of artificial intelligence in decision making in armed conflict", Humanitarian Law & Policy Blog, 24 October 2023, available online at https://blogs.icrc.org/law-and-policy/2023/10/24/algorithms-of-war-use-of-artificial-intelligence-decision-making-armed-conflict/.

20 Richard Eckel, "Microsoft Researchers' Algorithm Sets ImageNet Challenge Milestone", *Microsoft Research*, 10 February 2015 available online at https://www.microsoft.com/en-us/research/blog/microsoft-researchers-algorithm-sets-imagenet-challenge-milestone/.

21 "Computational Methods for Decision Making – Automated Image Understanding", *Office of Naval Research*, available online at https://www.nre.navy.mil/organization/departments/code-31/division-311/computational-methods-decision-making-automated-image.

AI DSS are also increasingly applied to enhance armed forces' ability to **collect and analyse data, and to generate information on the conflict environment**. For instance, a machine learning algorithm was reportedly used by governments to search the Global System for Mobile (GSM) communication metadata of millions of mobile phone users and to detect couriers carrying messages between non-state armed groups.[22] There are numerous research projects underway in this area. An example highlighted during the consultation was Project Maven, which focuses on computer vision – driven by machine learning and deep learning – to automate the processing, use and dissemination of massive amounts of data collected in operational areas across the globe, including to identify and classify objects of military interest.[23] There are also military research projects working on developing AI DSS for processing and analysing large, imperfect and unstructured data sets and providing visualization tools to allow users to analyse trends and glean value from the data.[24] This may include armed forces analysing the environment – including by examining the interconnectedness of objects or modelling the interiors of buildings, or assessing their own capabilities or those of friendly or adversary forces – by predicting the development of their respective capabilities or the adversary's behaviour. This information may then be used to advise armed forces' decisions about the place, time, person or object of an envisaged attack.

Finally, AI DSS may also be used in processes related to estimating expected **incidental civilian harm ("collateral damage") and weaponeering**. While in the past these assessments were done manually or by non-AI-based DSS, AI DSS are increasingly being used to assist human decision makers in these processes. For instance, AI DSS are used to aid human decision makers in collateral damage estimations by predicting the effects of explosives and other weapons.[25]

At the tactical level, AI DSS may provide human decision makers with **real-time information and recommendations** to inform decisions to act. For instance, virtual battle management systems use AI methods to process and analyse data to determine the optimal weapon for a specific target. Some of these systems are reportedly capable of providing recommendations based on parameters such as the target's location and the weapon's effectiveness, while taking into account the need to minimize collateral civilian harm and comply with the rules of engagement, in real time.[26] Similarly, technology has been developed to assist tactical commanders in estimating the position, strength and objectives of hostile forces and to predict their tactical movements in real time.[27] Finally, AI DSS are being touted as able to contribute to greater situational awareness across the battlefield by receiving and correlating information from national, strategic and tactical intelligence sensors and sources. An example highlighted during the expert discussions was an AI DSS, developed by a private company, that processes massive volumes of real-time data and presents them in a single view, with the stated purpose of enabling decision makers across roles and domains to make faster decisions, including those related to the use of force.[28]

22 Boulanin V and Verbruggen M, "Mapping the Development of Autonomy in Weapon Systems", *SIPRI*, 2017 available online at https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems.

23 John Keller, "Army to Brief Industry on Artificial Intelligence and Machine Learning for Intelligence Data Processing' *Military & Aerospace Electronics*", 4 October 2017, available online at https://www.militaryaerospace.com/computers/article/16726219/army-to-brief-industry-on-artificial-intelligence-and-machine-learning-for-intelligence-data-processing.

24 DARPA, "XDATA", available online at https://www.darpa.mil/program/xdata.

25 Frank Wolfe, "U.S. Collateral Damage Estimation Process Could Benefit from Real-World Data" Defense Daily, January 2021 available online at https://www.defensedaily.com/u-s-collateral-damage-estimation-process-benefit-real-world-data/weapons/.

26 Harry Lye, "IDF Acquires Rafael Fire Weaver Sensor-to-Shooter System" *Army Technology*, 3 February 2020 available online at https://www.army-technology.com/news/idf-rafael-weaver/ accessed 28 May 2021.

27 Michael Ownby and Alexander Kott, "Predicting Enemy's Actions Improves Commander Decision-Making", 2016 available online at http://arxiv.org/abs/1607.06759.

28 Palantir, "Gotham", available online at https://www.palantir.com/platforms/gotham/.

SECTION 3

ASSESSING THE CHALLENGES AND RISKS OF ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS USED IN MILITARY DECISION-MAKING ON THE USE OF FORCE

As discussed, one of the key drivers of the use of AI DSS applications in military decision-making on the use of force is the desire to accelerate and improve these decision-making processes, from the strategic level down to the tactical level. However, the use of these systems also gives rise to new challenges for military decision makers and potential risks for those affected by these decisions. While the military use of all DSS involves certain challenges, the application of increasingly complex AI-embedded DSS – especially those based on data-driven AI techniques – has created specific new challenges and risks. Many of these challenges mirror those relating to the application of AI in other contexts. However, they are likely to be exacerbated when AI is applied in a military context, owing to the specific nature of armed conflicts and military decision-making processes. This raises particular concerns, given the potential impact of these kinds of decisions on peoples' lives and dignity, and on communities.

Firstly, some of the risks relate to **AI systems' inability to perform abstract reasoning tasks, also known as the semantic gap**.[29] In contrast to human decision makers, an AI DSS may be capable of detecting the objects it has been trained to recognize in a technical sense, but it does not understand their meaning or context.[30] This lack of understanding leads to a greater risk of AI DSS "making mistakes" that humans would never make, for instance incorrectly labelling a person or an object. In the context of military decision-making on the use of force, the stakes involved in such labelling efforts are extremely high. In fact, the labelling effort may give rise to a range of IHL concerns, including the status of an individual under scrutiny, a key determinant for whether that person may be lawfully targeted.

Additionally, putting aside the fundamental concern that legal assessments under IHL must be made by humans, many key IHL rules regulating the use of force presume the application of evaluative decisions and value judgements, such as the presumption of civilian status in case of "doubt", or the assessment of what constitutes "excessive" expected incidental civilian harm in relation to anticipated concrete and direct military advantage. While these judgements certainly involve great challenges for human decision makers, they are arguably not suitable for encoding in AI DSS machine processes. From a technical point of view, it has been stressed that there is no way to arbitrate between irreconcilable, conflicting definitions of these highly qualitative and contextually variable notions using purely computational or mathematical means. Ethically, allowing AI DSS to proceed with such evaluative legal judgements may have an irreversible impact on peoples' lives and risks undermining their fundamental dignity.

29 International Committee of the Red Cross, "Autonomy, artificial intelligence and robotics: Technical aspects of human control", ICRC, 20 August 2019, p. 20.

30 Andrew D Selbst and others, "Fairness and Abstraction in Sociotechnical Systems", Proceedings of the Conference on Fairness, Accountability, and Transparency, 59, 2019, pp. 5–6.

Moreover, **AI DSS outputs are not free of bias**.[31] Considering that AI DSS are used to classify, rank and rate, and to produce various kinds of useful output to inform military decisions on the use of force, they will inevitably need to discriminate – in the technical sense of making distinctions between people or objects based on certain features. However, in contrast to existing human biases, the use of AI DSS has the potential to not only reinforce existing biases but also introduce new ones – potentially unknown to the developer – on a large scale. This can be seen in various civilian applications of AI DSS, which have amplified the resulting negative impacts.[32]

Closely related to these biases is the fact that AI DSS based on machine learning and deep learning models highly depend on the quality and quantity of their training data.[33] However, in armed conflicts it is very difficult to obtain such high-quality, representative training data for a specific military decision-making task. Armed conflicts often involve novel and context-specific aspects that do not lend themselves to producing transferable, repeatable data. Furthermore, in such contested environments, ensuring an element of surprise over the adversary is considered key to victory. In other words, inherent uncertainty and unpredictable situations constitute a characteristic of warfare. The resulting lack of training data in such contexts poses serious concerns related to the use of AI DSS to predict risk, as the systems outputs in these cases tend to be based on past data. Ultimately, this is likely to impact the adequacy and utility of an AI DSS in military decision-making on the use of force in these kinds of contexts.

Beyond that, using AI DSS to predict the risk posed by the actions of certain individuals or groups may also be legally problematic, as the output of the AI DSS is not based on the (future) behaviour of the individual under scrutiny, but on the (past) behaviours of others. In fact, the lawfulness of using recommendations produced by such systems in military decision-making on the use of force may be called into question, since IHL requires that a party to an armed conflict may only target a person based on his or her individual activities or status as a member of an armed force (and not on the activities of others). Additionally, in some instances, the collection of the required training data raises serious privacy concerns.[34]

Moreover, data-driven AI DSS are particularly vulnerable to adversarial attacks. In the civilian sector, there are abundant examples of how AI DSS used for image classification can be misled.[35] Armed conflict is characterized by belligerents trying to gain an advantage over one another. As a result, adversarial attacks on AI DSS become particularly likely in this context. Indeed "ruses", or methods of deception, are lawful as long as they do not amount to perfidy or lead to the attacking force mistakenly directing attacks against civilians or civilian objects.

Further concerns relate to the lack of understandability and predictability of AI systems. As documented in relation to various applications, the characteristics of current AI systems drastically diminish the human ability to know *what output the system will produce and why the system will produce a given output*. Hence, it becomes more difficult for human decision makers to properly evaluate the quality of AI DSS outputs and solve potential conflicts between their own opinion and the AI DSS suggestion.[36] In any case, AI DSS outputs supporting decisions on the use of force will need to be assessed as simply one component of the information available from all sources,

31 See: AIAAIC, "AIAAIC Repository", available online at https://www.aiaaic.org/aiaaic-repository. AIAAIC catalogues such cases since 2019 based on open-source data.

32 *Ibid*.

33 European Union Agency for Fundamental Rights, 'Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights', 2019, at 4–5, available online at https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect.

34 Robin Geiss and Henning Lahmann, "Protection of Data in Armed Conflict", 97 *International Law Studies*, 2021, at 8; Marko Milanovic, "Human Rights Treaties and Foreign Surveillance: Privacy in the Digital Age", Social Science Research Network, 2014, at 1.

35 Chenwei Li and others, "Misleading Image Classification with Multi-shearing and Random Padding": In Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering (EITCE '22), *Association for Computing Machinery*, New York, NY, USA.

36 Greipl, *supra* note 5.

as part of the obligation to take all feasible precautions to verify that targets are legitimate military objectives, and to avoid – or at least minimize – incidental harm to civilians.

This is also relevant to the testing of such systems before deployment, and the retesting or approval of AI DSS whose functioning may change during use. In the context of life and death decisions, these AI characteristics increase the risk of human decision makers, owing to reliance on erroneous AI DSS outputs, making inappropriate decisions, which can have significant consequences and potentially result in IHL violations, as well as conflict escalation.

This risk is further exacerbated by the **human tendency to rely on AI DSS outputs, even in the face of other, conflicting information that would support an alternative outcome or decision**, also known as **AI over-reliance**. Experts have highlighted several factors explaining why and when humans tend to rely too heavily on AI DSS outputs, including automation bias, humans' lack of technical understanding of the AI DSS, and situations of stress and pressure. The latter is a particularly worrying factor in military decision-making on the use of force, considering the growing tempo of military operations and the drive to use AI DSS to ensure faster decision-making. There is a concern that, in such contexts, military decision makers would not only be subject to a high level of pressure, but in some situations would be left with almost no time to assess, cross check or challenge AI DSS outputs. As a result, there is a risk that human decision makers may become passive supervisors of AI DSS systems outputs, rather than active controllers, and therefore blindly act on these outputs. The closer these decision-making tasks get to the actual use of force, the more human over-reliance may result in situations where the functions of target selection and the application of force are – in effect – being carried out without any real human intervention (raising questions regarding the degree to which this process really differs from an AWS, where targets are selected and force is applied without human intervention).

Finally, concerns arise regarding the use of AI DSS that continuously modify or update their parameters to improve their performance through online learning – as opposed to systems that stop "learning" after the development phase. While such systems could offer significant gains by progressively "fitting" their statistical models to the deployed environment, thus potentially reducing brittleness, such systems could also potentially acquire behaviours that have not been tested and certified, including emergent behaviours that may not be foreseen by the developers.

In conclusion, as this preliminary assessment shows, some of the inevitable failures or dysfunctions involving the use of AI DSS come with significant risks in the context of military decision-making on the use of force, due to the implications of such decisions for peoples' lives and dignity. Use of AI DSS may enable humans to make better decisions during hostilities, in compliance with IHL, and minimize risks for civilians by facilitating quicker and more widespread collection and analysis of available information. However, over-reliance on these algorithmically generated analyses or predictions might also lead to worse decisions or to violations of IHL, and exacerbate risks for civilians, especially given the current limitations of the technology.[37]

These findings stress the importance of an in-depth understanding of the capabilities and limitations of AI DSS, as well as the risks associated with human-AI DSS interaction in such critical decision-making processes. Furthermore, the findings suggest a need to develop specific measures that will preserve human judgement in military decision-making on the use of force to reduce humanitarian risks, mitigate ethical concerns and ensure compliance with IHL. The question of how to do so, given the use of increasingly complex AI DSS, is addressed in the next section.

37 ICRC *supra* note 1.

SECTION 4

INITIAL ASSESSMENT OF POSSIBLE SOLUTIONS: ADDRESSING THE CHALLENGES AND RISKS OF ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS IN MILITARY DECISION-MAKING ON THE USE OF FORCE

How can humans utilize AI DSS outputs in military decision-making on the use of force while effectively preserving human judgement in legal decisions? In identifying potential solutions, two interrelated questions should be kept in mind: to what extent can some of the existing challenges relating to human-AI DSS interaction realistically be reduced, and what constraints should be placed on the use of AI DSS in military decision-making on the use of force, given that some of the challenges and risks identified will likely persist?

Taking this into consideration, along with the risks and challenges presented in section 3, this report draws attention to a non-exhaustive list of possible solutions, divided into two categories: technical solutions and those that allow humans to critically engage with and use AI DSS and their outputs, taking into account the features of human-machine interaction.

EXPLORING TECHNICAL SOLUTIONS

Although no viable, measurable criteria for grading understandability and predictability currently exist, certain levels are considered essential for humans to effectively utilize AI DSS outputs. Those levels may arguably need to be higher in decision-making processes with a potential impact on peoples' lives and dignity, including decisions on the use of force.

The importance of taking these considerations into account prior to the implementation of AI DSS has been emphasized. Considering that it is possible to anticipate and understand AI DSS outputs and their associated risks only when they operate in conditions for which they have been specifically tested, it is crucial to take these considerations into account during the development, testing and evaluation phases. Technical solutions could include training AI DSS on data sets that cover the broadest possible range of inputs, to reduce the likelihood of AI DSS encountering unknown or unforeseen inputs, once deployed.

However, given the inherent operational unpredictability of complex and dynamic environments such as armed conflicts, it is highly doubtful whether developers and their organizations would be able to certify that a system would respond safely or appropriately to any input or conditions that it might encounter. In the light of this, the use of AI DSS should at least be limited to contexts for which they have been specifically and rigorously tested. For example, a computer vision system that has only been trained and tested to recognize tanks during clear-weather and daytime operations should not be used in night-time operations or in adverse weather

conditions.[38] Likewise, the use of AI DSS with continuous online learning functions requires greater caution due to their increased unpredictability (see section 3). Accordingly, armed forces may need to adopt measures following the deployment of an AI DSS, such as continuous tracking and performance monitoring, as well as regular re-evaluations – including retraining, retesting and reapproval – of the system.

EXPLORING MEANS TO IMPROVE HUMAN DECISION MAKERS' ABILITY TO CRITICALLY ENGAGE WITH ARTIFICIAL INTELLIGENCE DECISION SUPPORT SYSTEMS OUTPUTS

Beyond these technical solutions, other measures may allow human decision makers to better engage with and use AI DSS and their outputs, taking into account the limitations of human-machine interaction in the decision-making process.

Confidence scores – widely used in the civilian sector– indicating the likelihood that an output is correct may be used to support human decision makers' ability to assess the quality of an AI DSS output. However, this measure is unlikely to suffice in itself, and may not be appropriate to ensure effective human judgement, especially in military decision-making on the use of force. For instance, while it might be tempting to believe that it facilitates the human decision makers' evaluation of the AI DSS output, it may potentially exacerbate problems relating to human over-reliance on the AI DSS output, including as a result of automation bias. Decision makers may end up blindly relying on a confidence score without understanding the technical implications of that score – or its limitations – and may not be willing or able to challenge or cross check that score.

As noted earlier, this risk is of particular concern in a military context, where decision makers may not currently be in a position to accurately assess the meaning of confidence scores and often operate under pressure, with little time available to question those scores.[39] Should confidence scores be used, it would arguably be necessary to at least provide decision makers with general guidelines, emphasizing that these scores are mere technical outputs that cannot be equated with legal judgements as to whether, for example, an object or person would be a lawful target. For example, military decision makers would have to understand that an AI DSS output indicating, with 95 per cent confidence, that a person is "in military uniform" is not equivalent to an IHL assessment of whether that person is a lawful target at that moment in time.

A further potential measure to enhance human decision makers' engagement with and appreciation of AI DSS outputs, despite existing understandability challenges, is the use of AI DSS "model cards". In the civilian sector, such model cards were developed to enable experts and non-experts alike to obtain "simple overviews of models' ideal forms of input, visualize some of their key limitations, and present basic performance metrics".[40] In contrast to their use in the civilian sector, the applicability of these measures in a military context would be limited because the model cards might be used by adversaries to understand and exploit AI DSS vulnerabilities.

Developers may also play a key role, not only in the system and interface design, but by providing clear and useable guidance on the use of AI DSS. Given their technical knowledge, developers

38 Paul Maxwell, "Artificial Intelligence Is the Future of Warfare (Just Not in the Way You Think)", *Modern War Institute*, 20 April 2020 available online at https://mwi.westpoint.edu/artificial-intelligence-future-warfare-just-not-way-think/.

39 Erik Davis, "The Need to Train Data-Literate U.S. Army Commanders", War on the Rocks, 17 October 2023 available online at https://warontherocks.com/2023/10/the-need-to-train-data-literate-u-s-army-commanders/.

40 Google, "Model Cards: The value of a shared understanding of AI models", available at https://modelcards.withgoogle.com/about.

are well placed to help users better understand how these systems operate in a specific context. Accordingly, developers could offer practical guidance to decision makers on the system's performance, including on its potential functions and the environments in which it will potentially be deployed, highlighting its capabilities and limitations, modelling potential failures and their impacts, and drawing attention to any other potential harms that could arise from its use. Such practical guidance on AI DSS capabilities and limitations offers a potential means for military decision makers to better anticipate the risks of using AI DSS systems, and more reliably assess the compliance of their decisions with IHL.

At the same time, efforts to enhance human decision makers' ability to effectively utilize AI DSS outputs will require time and resources ahead of the implementation of these systems. Thus, it will be important to train AI DSS users.

Finally, a decision maker's practical ability to effectively utilize AI DSS outputs will also depend on the speed of present and future processes for making decisions on the use of force. While the underlying rationale for using AI DSS is to accelerate decision-making, it has been suggested that there may be a need to impose "speed limitations" at certain points to ensure sufficient time and space for human deliberation. Even if faster decision-making is considered a major military advantage, that advantage can only be ensured if humans retain the ability to interact effectively with AI DSS in these decision-making processes in a way that allows them to comply with their legal obligations in carrying out attacks. Similarly, the ICRC has suggested that "in order to preserve human judgement, systems may need to be designed and used to inform decision-making at "human speed", rather than accelerating decisions to "machine speed" and beyond human intervention".[41]

41 ICRC, *supra* note 1; See also *supra* note 20.

SECTION 5

MAIN FINDINGS AND POSSIBLE NEXT STEPS

Armed forces are exploring the use of increasingly complex AI techniques enabling AI DSS to support more and more complicated decision-making tasks related to the use of force. While these applications offer new opportunities, they inevitably give rise to new challenges and risks. There is likely no easy solution to these challenges and risks in all eventualities, but the expert consultations conducted to prepare this report have shown a need to further explore practical measures to preserve effective human judgement in such critical decisions. Based on an expert consultation with a diverse group of multidisciplinary experts, we identified the following main takeaways:

The introduction of AI to DSS for military decision-making on the use of force in armed conflicts adds a new dimension to existing challenges relating to non-AI-based DSS. The use of these systems raises new questions regarding the understandability and predictability of their outputs, the speed at which they operate, and barriers to the human ability to effectively assess the accuracy of their outputs. As a result, the use of AI DSS capabilities has the potential to reduce the human judgement involved in military decision-making on the use of force in armed conflicts, thus raising humanitarian, legal and ethical questions.

The military application of AI DSS and of AWS raises distinct legal and conceptual issues, even though some of the underlying technology may be very similar. For instance, some of the applications of AI DSS, such as those developed for automated target recognition, could form part of an AWS. However, once a weapon system carries out the processes of selecting and applying force to a target without human intervention (i.e. an AWS), distinct humanitarian, legal and ethical concerns arise. This has prompted the ICRC to recommend a specific regulatory response in the form of new legally binding rules on AWS.[42]

The realities of warfare mean that the challenges and risks linked to the use of AI DSS in other contexts will likely be exacerbated when AI DSS are used in military decision-making on the use of force in armed conflicts. This raises particular concerns, given the significant potential impact of these kinds of decisions on peoples' lives and dignity, and on communities.

Preserving human judgement in military decision-making on the use of force in armed conflicts is crucial to reducing humanitarian risks, addressing ethical concerns and facilitating compliance with IHL. This may require new approaches to existing challenges arising from the interaction between humans and AI DSS, as well as technical requirements relating to the use AI DSS. Some approaches may help to address existing technical challenges (e.g. predictability, understandability, and bias), while others may help to improve human decision makers' ability to critically engage with and use AI DSS outputs (e.g. mitigating automation bias). Nevertheless, to ensure the viability of such approaches, it is important to assess them in relation to their use in the specific context of an armed conflict, and to require that their use is restricted to tasks and contexts for which they have been specifically and rigorously tested.

42 International Committee of the Red Cross, "ICRC position on autonomous weapon systems", 12 May 2021 available online at https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems.

The use of AI DSS in military decision-making on the use of force may require additional measures and constraints to mitigate risks for people affected by armed conflict, and to facilitate compliance with IHL. Many of the existing challenges of human–machine and human–AI interaction are likely to persist, while certain technical limitations may be insurmountable. Therefore, it may be necessary to place certain constraints on the use of AI DSS in decisions on the use of force. These may include: restricting the use of AI DSS to certain tasks or decisions and/or to certain contexts; placing specific constraints on the use of AI DSS with continuous learning functions, due to their more unpredictable nature; and slowing down the military decision-making process at certain points to allow humans to undertake the qualitative assessments required under IHL in the context of specific attacks.

THE WAY FORWARD

The expert consultations conducted as part of this project revealed a need for further research and dialogue in this area, in order to better understand the measures and constraints that may be required with regard to the design and use of AI DSS to mitigate the risk of harm to people affected by armed conflict, and to ensure compliance with IHL. Further analysis will be needed to identify applications of AI DSS in this context that have the biggest impact on decisions on the use of force.

ANNEX 1

BACKGROUND DOCUMENT

DIGITALIZATION OF CONFLICT: HUMANITARIAN IMPACT AND LEGAL PROTECTION

New technologies have a profound impact on how wars are fought. While international humanitarian law (IHL) is applicable to all technological developments in warfare, the speed, scale, and transformative impact of today's extraordinary technological advances and the continuous merger of the physical and digital domains require a constant (re-)assessment whether new means and methods of warfare are compatible with existing IHL rules and whether IHL continues to provide the level of humanitarian protection it is meant to ensure in times of armed conflict.

The project "**Digitalization of Conflict: Humanitarian Impact and Legal Protection**", a joint initiative between the **International Committee of the Red Cross (ICRC)** and the **Geneva Academy of International Humanitarian Law and Human Rights**, aims to explore humanitarian consequences and protection needs caused by the digitalization of armed conflicts and the extent to which these needs are addressed by international law, especially IHL.

The digitalization of armed conflict is a dynamic process that encompasses the increasing use of digital means and methods of warfare based on a range of rapidly evolving technological developments, most notably in cyber technologies, artificial intelligence (AI), machine learning, sensor systems, and robotics. The project considers the effects of these developments with a view to assessing the risks that they entail for conflict-affected populations and ensuring that the legal and policy framework provides adequate humanitarian protection in contemporary and future warfare.

The joint initiative adopts a multi-disciplinary perspective that takes into consideration the interrelated technical, military, ethical, policy, legal and humanitarian aspects to address three overarching questions:

1. *What risks, potential humanitarian consequences, and protection needs for conflict-affected populations arise on the digital battlefield?*
2. *Does international law, in particular IHL, adequately address these risks and protection needs?*
3. *If not, what recommendations could be developed in terms of law and policy beyond the existing IHL framework to mitigate these risks and address these protection needs?*

The first phase of the joint initiative explored questions related to the impact of cyber operations in armed conflict. In this **second phase, we will examine certain military applications of AI**.

SCOPE AND OBJECTIVE OF THE EXPERT WORKSHOP

This expert workshop will focus on the first of the joint initiatives' overarching questions, as it relates to **AI and related technologies in military decision-making on the use of force**.

Advancements in the field of AI, including machine learning, are likely to influence the way wars are fought. There have been extensive discussions on autonomous weapon systems, including the role of AI. However, other potentially significant applications from a humanitarian perspective are in cyber and information operations, and in military decision-making.[43]

Exploration of **AI-supported decision-making – or 'decision-support systems'** – for military operations is increasing, including for decisions that are governed by international humanitarian law. Many of the implications, including potential benefits and risks, remain underexplored and require further research and discussion.

This **workshop will focus specifically on the use of such technologies in military decisions to use force**, through (online) roundtable discussions with a range of experts with military, technical, policy and humanitarian expertise.

The ICRC and the Geneva Academy aim to **contribute to current analysis and understanding of these applications of AI in military decision-making, and their implications from a humanitarian perspective**. This will require an understanding of current and near-term military technological developments and what States and other armed actors might want to achieve, strategically, operationally, or tactically, through AI-supported decision-making in the use of force.

Against this backdrop, the workshop will be guided by three overall questions:

- *How are AI and related technologies currently being used in military decision-making on the use of force, and what are likely applications in the near-term (next 3-5 years)?*
- *How are these military applications likely to impact decisions on the use of force in armed conflict?*
- *What are the potential risks, benefits, humanitarian consequences, and protection needs for conflict-affected populations arising from these military applications?*

The **next section comprises brief background on decision support systems** in military decision-making on the use of force. It is authored by Arthur Holland Michel, an independent researcher and consultant for the ICRC for a current research project on this topic.

43 International Committee of the Red Cross, "Artificial intelligence and machine learning in armed conflict: A human-centred approach," *International Review of the Red Cross* 102 (913), Digital technologies and war, 2020, 463–479, available at: https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913.

BACKGROUND ON DECISION SUPPORT SYSTEMS IN MILITARY DECISION-MAKING ON THE USE OF FORCE

Arthur Holland Michel, Independent researcher and consultant for the ICRC

The process leading to the use of force by militaries is punctuated by many critical human decisions.[44] None of these decisions is easy.[45] Regardless of whether the process stretches across minutes or weeks, decision makers must account for a constellation of complex factors. These include evolving intelligence assessments and uncertainty[46] about the environment and the people in that environment – be they the civilian population, the adversary, and their own forces; the overarching strategic goals to which all the decision maker's actions must align; and the framework of legal, material, and operational restrictions to which any decision must conform. Taking all of these many variables into account, decision makers must seek to maximize the probability of achieving the objective with the lowest possible risk of adverse and/or unintended outcomes, while also – crucially – complying with all relevant IHL obligations, including taking all feasible precautions to avoid or at least minimize incidental harm to civilians.

Decision support systems (DSS) are computerized tools that are designed to aid such human decision-making. They do so by displaying, synthesizing or analysing relevant information, and/or by proposing options for how to achieve a goal. Even though DSS do not "make" decisions,[47] they directly and often significantly influence the decisions of humans decisionmakers.[48] As a result of advances in areas like computing, AI, data collection and communications, their capabilities will grow significantly in the years ahead, as will their influence on military decision making.

44 These include decisions related to understanding the environment, establishing an objective, developing a plan for how to achieve that objective, executing the action and evaluating the effects of the action once it has been completed. International Committee of the Red Cross, *Decision-Making Process in Military Combat Operations*, Geneva, 2013. An alternate formulation is that any process leading to a military action requires decision that answer the following three separate questions: "What is?", "What if?" and "What's next?". Gilles Desclaux, Baptiste Prebot, "Command and Control at the Autonomy and Cognitive Era: For a decision cycle augmented by the symbiosis between human and systems," 23rd International Command and Control Research and Technology Symposium, November 2018, Pensacola, United States.

45 Andreas Tolk and Dietmar Kunde, "Decision Support Systems – Technical Prerequisites and Military Requirements," 2000 Command and Control Research and Technology Symposium, June 2000, Monterey CA, United States.

46 Hasmik Atoyan, Jean-Marc Robert, Jean-Rémi Duquet, "Uncertainties in complex dynamic environments," *Journal d'Interaction Personne-Système*, Vol. 2, Num. 1, Art.5, January 2011; http://www.indiandefencereview.com/spotlights/uncertainty-and-risk-in-military-decision-making/

47 Interview with Svetlana Yanushkevich, November 2021. (All interviews were conducted online via Zoom unless otherwise noted); Interview with Peter Svenmarck, November 2021. Human decision-making is not solely based on mathematically defined criteria, parameters and goals: it also factors political, ethical, moral, emotional and strategic imperatives. In this regard, Decision Support Systems, which always support a human decision, are distinct from *Decision Systems* that make automated decisions. Marko Bohanec, *What is Decision Support?*, Jožef Stefan Institute, Ljubljana, 2001, p. 2

48 Merel Ekelhof, "Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting," *Naval War College Review*: Vol. 71: No. 3, Article 6, 2018, p. 23

DSS are considered to be helpful for enabling timely decisions that account for larger amounts of relevant information and reflect more mathematically optimal "solutions" to achieve a goal.[49] Compared to non-computerized methods for supporting a decision, DSS are regarded as being faster,[50] more comprehensive, more efficient, more consistent,[51] and less prone to errors.[52] Therefore, the ICRC has previously noted that such tools "may enable better decisions by humans in conducting hostilities in compliance with international humanitarian law [IHL] and minimizing risks for civilians by facilitating quicker and more widespread collection and analysis of available information."[53] In this way, they could potentially support the rigorous application of the law, in particular the rules of IHL, to the use of force – provided the intentions of the humans operating those systems are aligned with those norms.[54]

However, the ICRC has also observed that the "use and misuse" of DSS "could lead to increased risks for civilian populations."[55] DSS can and do fail, as can the people and processes that are supposed to ensure that their use does not result in decisions that have adverse and/or unintended outcomes. In some cases, these technologies and the people who use them have contributed to documented instances of undue harm in military operations. Therefore, an over-reliance on computerized analyses and predictions might "facilitate worse decisions or violations of international humanitarian law and" – likewise – "exacerbate risks for civilians."[56]

Preventing such harms could become more difficult in the years ahead. Thanks to the converging technological advances that are raising the profile of DSS in conflict, these systems are becoming more complex and will be used more widely to execute a greater range of functions. This growing complexity of DSS and their functions is likely to multiply the challenges of ensuring that humans make appropriate, contextually informed decisions on the basis of the DSS' computerized outputs. As a result, the *expanding* use of more *complex* decision support systems could reduce and *hinder* the application of critical legally required layers of human judgement in the process leading to the use of force, and thus shrink the *accountability surface*[57] of conflict. Modern machine learning, which has yet to be employed widely in critical DSS roles directly implicated

49 For a more comprehensive list of the specific perceived benefits or motivations for computerized DSS, see German Army Concepts and Capabilities Development Centre, *Artificial Intelligence in Land Forces*, Edition 2, Köln, 2019, p11; Development, Concepts and Doctrine Centre, *Joint Concept Note 2/17: Future of Command and Control*, United Kingdom Ministry of Defense, September 2017, pp.1–6

50 Interview with Margarita Konaev, October 2021; One study found that using a decision support tool called Integrated Course of Action Critiquing and Evaluation System (ICCES) for COA development reduced the time needed for a planning process from 16 hours down to 20 minutes. Robert Rasch, Alexander Kott and Kenneth D. Forbus, "Incorporating AI into military decision making: an experiment," *IEEE Intelligent Systems*, Volume: 18, Issue: 4, July–Aug. 2003

51 Merel Ekelhof, "Lifting the Fog of Targeting: 'Autonomous Weapons' and Human Control through the Lens of Military Targeting," *Naval War College Review*: Vol. 71: No. 3, Article 6, 2018 – p24; Walter A. Powell et al., "Results of an Experimental Exploration of Advanced Automated Geospatial Tools: Agility in Complex Planning," 14th International Command and Control Research and Technology Symposium, Washington, 15–17 June 2009.

52 DSS are also seen as a means to correct human cognitive biases that hamper decision-making, and to counteract the effects of factors such as lapses in concentration, fatigue, stress, or emotional state. Anonymous interview with an NGO employee, September 2021; Interview with Milind Kulshreshtha, September 2021; Cécile Godé and Jean-Fabrice Lebraty, "Improving decision making in extreme situations: The case of a military Decision Support System," The International Journal of Technology and Human Interaction, Vol. 9, N°2, 2013.

53 International Committee of the Red Cross, "Artificial intelligence and machine learning in armed conflict: A human-centred approach," *International Review of the Red Cross* 102 (913), Digital technologies and war, 2020, 463–479

54 Interview with Margarita Konaev, October 2021

55 International Committee of the Red Cross, "Artificial intelligence and machine learning in armed conflict: A human-centred approach," *International Review of the Red Cross* 102 (913), Digital technologies and war, 2020, 463–479.

56 *Ibid*.

57 That is, the degree to which *human* agents can be held accountable for harms.

in the process leading to the use of force,[58] is likely to pose additional challenges in this regard – especially with regard to bias, predictability and understandability. These challenges could be a particular concern in the use of DSS at tactical levels, close to the application of force itself, and in complex scenarios where the time available for human decision-making is highly constrained.

58 Interview with Peter Svenmarck, November 2021; Priya Narayanan et al., *First-Year Report of ARL Director's Strategic Initiative (FY20-23): Artificial Intelligence (AI) for Command and Control (C2) of Multi-Domain Operations (MDO)*, DEVCOM Army Research Laboratory, Adelphi May 2021 – p. 3; Only in the autumn of 2021 did the U.S. military announce that it had employed a machine-learning based decision support tool (a target recognition system) for the first time in a kill-chain that resulted in a kinetic strike. Amanda Miller, AI Algorithms Deployed in Kill Chain Target Recognition, *Air Force Magazine*, 21 September 2021, https://www.airforcemag.com/ai-algorithms-deployed-in-kill-chain-target-recognition/.

ANNEX 2

BACKGROUND DOCUMENT

INVITED EXPERTS

- **Rachel AZAFRANI**, AI and IoT Security Strategist, Digital Diplomacy, Microsoft
- **Ingvild BODE**, Associate Professor, Department of Politics and Public Administration, University of Southern Denmark (SDU)
- **Vincent BOULANIN**, Senior Researcher, Stockholm International Peace Research Institute (SIPRI)
- **Bérénice BOUTIN**, Senior Researcher and Project Leader, Designing International Law and Ethics into Military Artificial Intelligence (DILEMA), Asser Institute
- **Jason BROWN**, Professional Services Manager – Public Sector, Google
- **Raja CHATILA**, Professor Emeritus, Sorbonne University
- **Anthony CORSO**, Executive Director, Stanford Center for AI Safety, Stanford University
- **François-Régis DABAS**, Head, Conventional Weapons and Military Space Activities, Arms Control Division, French Army Staff
- **Tessa DE HAAN**, Senior Policy Officer, New Technologies and Arms Control, Ministry of Foreign Affairs, The Netherlands
- **S Kate DEVITT**, Chief Scientist, Trusted Autonomous Systems Defence Cooperative Research Centre, Australia
- **Alexandra DUCA**, Ministry of Defence, Austria
- **Jonathan ELLIOTT**, Chief of Testing and Evaluation, Chief Digital and Artificial Intelligence Office (CDAO), Department of Defence, USA
- **Eugenio GARCIA**, Deputy Consul General, Head of Science, Technology and Innovation, Consulate General of Brazil in San Francisco
- **Matthew HARDING**, Director Legal Services (Operations and Intelligence Law), New Zealand Defence Forces
- **Arthur HOLLAND MICHEL**, Independent researcher, and consultant for the ICRC on AI and related technologies in military decision-making on the use of force
- **Margarita KONAEV**, Research Fellow, Center for Security and Emerging Technology (CSET), Georgetown University
- **Quentin LADETTO**, Research Director – Technology Foresight, Armasuisse, Switzerland
- **Kobi LEINS**, Visiting Honorary Research Fellow, Centre for Science and Security Studies, Department of War Studies, King's College London
- **Dustin LEWIS**, Research Director, Harvard Law School Program on International Law and Armed Conflict (HLS PILAC)
- **Jennifer MCKEEN**, Directorate of International Law, Canadian Armed Forces
- **Elisa MO**, Deputy Director, Future Systems and Technology Directorate, Ministry of Defence, Singapore
- **Simon MONCKTON**, Defence Research and Development Canada (DRDC)
- **Brian MULLINS**, CEO, Mindfoundry
- **Benjamin NG**, Senior Analyst (Emerging Technologies), Defence Policy Office, Ministry of Defence, Singapore
- **Ravi S PANWAR**, Distinguished Fellow, The United Service Institution of India
- **Frank SAUER**, Senior Research Fellow, Bundeswehr University Munich
- **Michael SIEGRIST**, Legal Adviser, Federal Department of Foreign Affairs, Switzerland
- **Jacob SHAHA**, US Army Artificial Intelligence Integration Center
- **Mariarosaria TADDEO**, Associate Professor and Senior Research Fellow, University of Oxford

- **Wichuta TEERATANABODEE**, Senior Analyst, Military Transformations Programme, S. Rajaratnam School of International Studies (RSIS)
- **Catherine TESSIER**, Research Director, Department of Information Processing and Systems, National Office for Aerospace Studies and Research (ONERA), France
- **Hideo TOMIKAWA**, Senior Fellow, National Institute for Defence Studies, Japan
- **Alexandria WALDEN**, Global Head of Human Rights, Google
- **Caroline WÖRGÖTTER**, Ministry of Foreign Affairs, Austria
- **XU ZHIDONG**, Associate Professor, National Defence University, China
- **ZENG YI**, Professor, Institute of Automation, Chinese Academy of Sciences; Chief Scientist in AI Ethics and Governance, Institute of AI International Governance, Tsinghua University

ICRC AND GENEVA ACADEMY

- **Maya BREHM**, Legal Adviser, Arms and Conduct of Hostilities Unit, ICRC
- **Neil DAVISON**, Senior Scientific and Policy Adviser, Arms and Conduct of Hostilities Unit, ICRC
- **Laurent GISEL**, Head, Arms and Conduct of Hostilities Unit, ICRC
- **Jonathan HOROWITZ**, Legal Adviser, ICRC Delegation for the USA and Canada
- **Abhimanyu JAIN**, Legal Adviser, Arms and Conduct of Hostilities Unit, ICRC
- **Chiara REDAELLI**, Research Fellow, Geneva Academy
- **Marco ROSCINI**, Swiss Chair of International Humanitarian Law, Geneva Academy
- **Ruben STEWART**, Thematic Adviser, Armed Forces and Security Unit, ICRC
- **Mauro VIGNATI**, Adviser on New Digital Technologies of Warfare, Arms and Conduct of Hostilities Unit, ICRC

MISSION

The International Committee of the Red Cross (ICRC) is an impartial, neutral and independent organization whose exclusively humanitarian mission is to protect the lives and dignity of victims of armed conflict and other situations of violence and to provide them with assistance. The ICRC also endeavours to prevent suffering by promoting and strengthening humanitarian law and universal humanitarian principles. Established in 1863, the ICRC is at the origin of the Geneva Conventions and the International Red Cross and Red Crescent Movement. It directs and coordinates the international activities conducted by the Movement in armed conflicts and other situations of violence.

facebook.com/icrc
twitter.com/icrc
instagram.com/icrc

International Committee of the Red Cross
19, avenue de la Paix
1202 Geneva, Switzerland
T +41 22 734 60 01
shop.icrc.org

4758/002 02.2024