

DECISIONS, DECISIONS, DECISIONS: COMPUTATION AND ARTIFICIAL INTELLIGENCE IN MILITARY DECISION-MAKING



ICRC OBSERVATIONS ON EXTERNAL REPORT

ICRC OBSERVATIONS ON EXTERNAL REPORT

Military applications of artificial intelligence (AI) and related data collection and analysis technologies have significant implications in armed conflict. Some of the most far-reaching relate to the use of these technologies to support decision-making in conflict. Among these are ‘decision support systems’ (DSS), computerized tools that provide outputs (i.e. analyses, recommendations and predictions) to inform human decision-making. Possible military uses of DSS are broad, from supporting decisions about who, what or where to attack, to decisions on who to detain and for how long, or recommendations on military strategy and specific operations, including attempts to predict or pre-empt adversaries’ actions.¹

The expanded development and use of machine learning-based DSS is seen by militaries as one of the most immediately useable options for AI. This is part of military efforts to analyse large amounts of information in order to improve their understanding of the operational environment and shorten military decision-making times, particularly the time between identifying a target and taking action against it. Military applications of increasingly machine learning-based DSS may therefore have a significant impact on decisions to use force, including the process by which such decisions are made and the consequences that may result.

Some argue that the use of these DSS can support human decision-making in a way that facilitates compliance with international humanitarian law (IHL) and minimizes the risks for civilians, for example, by enabling quicker, more widespread and more complex analysis of relevant information in a given situation. Indeed, IHL requires that all feasible precautions be taken to avoid civilian harm when conducting hostilities and calls for decisions to be reached based on an assessment of all information from available sources.² Others caution, for example, that overreliance on AI-generated analyses, recommendations and predictions poses concerns for civilian protection and compliance with international law, especially given the opaque, unexplainable and biased nature of many of today’s machine learning techniques.

To better understand these developments, the International Committee of the Red Cross (ICRC) commissioned the report by Arthur Holland Michel: *Decisions, Decisions, Decision: Computation and Artificial Intelligence in Military Decision-Making*.³ The author draws on interviews with experts and desk research to highlight some major trends in military DSS development, identify key aspects of machine-learning based DSS in particular and discuss salient implications for military decision-making on the use of force.

Based on the findings of this research, the ICRC has identified several issues that deserve further attention in considering the implications of these military technology developments.

¹ ICRC, *ICRC Position Paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach*, International Review of the Red Cross (IRRC), No. 913, March 2021: <https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913>.

² ICRC, Customary IHL Rules 14, 15 and 17.

³ The enclosed report is the work of the author and does not necessarily represent the views or position of the ICRC.

LEGAL DETERMINATIONS MUST BE MADE BY HUMANS

Under IHL, humans must make legal determinations, such as whether the expected incidental civilian harm from an attack will be excessive in relation to the concrete and direct military advantage anticipated.⁴ And the individual – along with their commanders and other superiors – will bear responsibility for these decisions. This is particularly relevant for those holding command responsibility. In reaching their determinations, the people responsible for planning, deciding upon or executing attacks might take DSS output into account, for example, from the Automatic Target Recognition, Weapons Effects Modelling or Collateral Damage Estimation systems described in the report. However, technical indicators can never substitute human, legally mandated judgements, such as whether a person or object can lawfully be targeted. The responsibility and the accountability for those legal determinations rest with individuals and their commanders and cannot be transferred to a machine or computer program; it is humans who must comply with IHL and accordingly, humans are required to exercise judgement regarding the lawfulness of an attack. At the same time, the way in which DSS function and how human users interact with their output may raise challenges for accountability processes, especially when it comes to enforcing individual criminal responsibility.

The following sections examine how the human judgement required by IHL may be affected or impeded by the military use of increasingly complex machine learning-based DSS. They highlight design and operational considerations that should be factored in by states and others developing and using DSS to ensure humans remain responsible for decisions in the conduct of hostilities.

MACHINE LEARNING IN DSS: INCREASED COMPLEXITY, INCREASED POINTS OF FAILURE

Machine learning is seen by many militaries and companies developing and providing DSS for use in armed conflict as a key enabling technology that will fuse and translate sensor data into descriptive analyses and prescriptive courses of action or predictions that underpin decisions on the use of force. Developers' hopes are that these types of DSS will offer more complex assessments and nuanced outputs to improve situational awareness for decision makers, thereby offering a military advantage and the potential to be used to help ensure decision-making is in line with legal obligations.

The report identifies various interacting uncertainties, assumptions and biases that must be accounted for in the technical functioning of decision support tools, in the human user and in the interaction between the two. Many of these are particular to, or accentuated by, increasingly complex DSS that incorporate machine learning.

SYSTEM LIMITATIONS

In terms of the system itself: (1) machine learning-based DSS introduce heightened challenges with respect to predictability, understandability and explainability for the user in how and on what basis such DSS produce their analyses, recommendations and predictions; (2) the more information that the DSS extracts from data and the more complex the planning that it embeds, the greater the number of assumptions on which the DSS is based and the greater the risk that the output exacerbates or introduces bias; (3) machine learning systems have a higher propensity to fail in unpredictable ways and to create error modes that are impossible to anticipate in testing; and (4) adversary forces may behave in deliberately unpredictable and deceptive ways

⁴ ICRC, [2019 Challenges Report](#): "... legal obligations under IHL rules on the conduct of hostilities must be fulfilled by those persons who plan, decide on, and carry out military operations. It is humans, not machines, that comply with and implement these rules, and it is humans who can be held accountable for violations. Whatever the machine, computer program, or weapon system used, individuals and parties to conflicts remain responsible for their effects."

that quickly render training data obsolete.

This added complexity can be contrasted with more “traditional” DSS, which are based on deterministic (rule-based) algorithms that will always produce the same output for a given input. This makes the outputs more predictable, explainable and understandable, but it can limit their utility for complex assessments with unknown variables that cannot be encoded.

By contrast, machine learning-based DSS work on a stochastic (probability) basis. Rather than using predefined criteria, such a system might, for example, identify an object like a missile based on the degree to which it matches the other objects that are labelled as “missiles” in that system’s training data. Having built their own rules (or model) for a given problem based on training data, these DSS tools will match observed characteristics in sensor or other data against that model. All DSS are built around assumptions that assign meaning to mathematically definable attributes. Rather than being scripted (coded) by humans, in machine learning-based DSS, these assumptions can be scripted by the system itself.

As the report highlights, machine learning therefore brings novel challenges in terms of predictability, explainability and understandability for the user in how and on what basis DSS produce their analyses, recommendations and predictions.

These challenges will likely be exacerbated by the difficulty of ensuring that data used for training, testing and verifying machine learning-based DSS are sufficiently representative of the data the DSS will process during its use to produce a particular analysis or output. This is something that militaries developing and using these systems will need to address.

Further, as Holland Michel argues, DSS that are used for more complex tasks embed a larger number of assumptions, which “could result in unintended or unlawful harm and diverge from the wishes of those developing or deploying them.” For example, the output could exacerbate well-documented bias that would fail to account for the realistic presence, activities and risks certain civilians face and their actions and reactions during armed conflict, thereby placing them at greater risk (such as individuals or groups of a certain age and gender, people with disabilities or people carrying weapons legitimately, such as law enforcement officials).

With increasingly complex machine learning-based DSS, it becomes harder to make these assumptions available to the user and to validate them at either the time of development, in testing or during use.

This is not the only hurdle that must be overcome in order to conduct sufficient and effective testing, evaluation, validation and verification of these systems. Such processes would also need to be able to account for the higher propensity of machine learning-based systems to “fail in unpredictable ways” and to create “error modes that are impossible to anticipate in testing”, but which might create “cascading errors” that perpetuate through all subsequent functions. This is one of the unique, additional uncertainties introduced by the advent of machine learning-based DSS.

These errors might also arise due to hacking or spoofing, since models developed using machine learning are known to be vulnerable to “adversarial techniques that cause the system to generate unpredictable erroneous outputs, often in a manner that is undetectable to human operators.” Unlike in peacetime, these adversarial actions must always be expected in conflict, as should deception tactics such as ruses of war. Military domains are marked by imperfect information, shifting dynamics and adversary forces who will behave in deliberately unpredictable and deceptive ways that quickly render training data obsolete.

HUMAN MACHINE INTERACTION CHALLENGES

The technical aspects of the system outlined above will clearly pose challenges for the human user interacting with the DSS. Overall, Holland Michel's analysis suggests that, with the increasing complexity of machine learning-based DSS, the user may be willing, yet simply unable to engage meaningfully with the output. It becomes impossible for them to account for the system's limitations or to detect errors due to the "uncertainties, assumptions and biases that are embedded in DSS outputs – especially machine learning based systems – in relation to the unique context of the decision it supports."

THE 'RUBBER STAMP' TRAP AND AUTOMATION BIAS

Further affecting the degree to which humans can meaningfully engage with decision-making based on DSS outputs, the report argues, is the way in which one complex DSS can process several different analyses into a single output for the user. This consolidation of many distinct decision-making steps into one data and machine learning-driven output, when previously each step had human involvement, could automate an entire decision-making chain, "such that the human's only role is reduced to either approving or negating a proposed plan for the use of force". The report points out that this could reduce users' decision to a single judgement of whether or not to rely on the system. And as the number of roles merged by the system increases, so too does the complexity of that decision on whether or not to "trust" the system.

The resulting risks may be exacerbated by the human-machine interaction phenomenon of automation bias. Automation bias, or over trust, can result in complacency and overreliance on the machine output, especially if the output fits with the users' expectations.

One overall risk, therefore, is an overreliance on machine outputs as the basis for legal decisions and ethical determinations that require context-specific and value-based judgement by humans.

To address this risk, there is a need to ensure that the human decision maker is not blindly relying upon DSS output and simply providing a human 'rubber stamp'. To do this, human users must be in a position to exercise sufficient independent scrutiny of the available information, including accounting for the capabilities and limitations – such as assumptions, biases and uncertainties – of the DSS in the circumstances of its use, while retaining the operational capacity to disregard outputs as appropriate. Careful consideration will be needed in determining the processes, constraints and training that enables users to actually make the assessments required by law in decisions on the conduct of hostilities.

For the use of a DSS to support lawful decisions on the use of force, the user must have some means of assessing its limitations and capabilities in the specific context of use. However, the report argues that there are "inherent obstacles" for users to do this, and these obstacles increase "as DSS become more complex and as they are used for a wider variety of less mathematically definable tasks." This may also require drastically re-thinking the way human-machine interaction challenges are assessed as compared to "traditional" rule-based DSS, and consequently adapting the operators' and commanders' training in how to use them. Such training will need to account for the diversity of situations in which a system might be used, the unpredictable nature of warfare and the ruses and other deception techniques that any adversary might be expected to use.

PRESERVING SUFFICIENT TIME AND SPACE FOR HUMAN DELIBERATION

The report identifies the increased speed of decision-making as a prominent military driver for developing machine learning-based DSS. Such systems may facilitate this by reducing the volume of information decision makers need to assess and integrating ordinarily separate analytical activities into a single tool. According to Holland Michel, “By some estimates, a target search, recognition and analysis activity that previously took hours could be reduced to minutes, and a process that previously took minutes could potentially be reduced to seconds.”

The expectation is that machine learning-based DSS will fuse and analyse larger and more diverse data sets, while also expanding from descriptive analyses (e.g. ‘vehicle is classified as an armoured fighting vehicle’) to include increasingly prescriptive and predictive inferences and proposals for courses of action (e.g. ‘85% confidence prediction that the vehicle is an enemy battle tank which poses a threat, your best course of action is striking with long range artillery’). The report notes that militaries are developing these systems with the hope they will lead to better threat detection, including assessment of objects and events based on limited information, resulting in improvements in ‘situational awareness’.

The report also highlights a trend towards DSS use “at the edge”, in other words, DSS being used further down the chain of command in tactical operations and dynamic targeting. DSS may also enable real-time analytics and adaptability to the environment, given that machine learning-based DSS can be continuously updated during use.

However, while an increase in the tempo of decision-making may present a military benefit, it can create additional risks to civilians if time pressures prevent thorough analysis and consideration of available information from different sources. Indeed, reducing tempo to allow for ‘tactical patience’ has been recognized as a technique to reduce civilian casualties.⁵ This type of time compression in decision-making will also be relevant at the strategic level, where conflicts can be avoided or escalated depending on how users respond to machine-generated analyses. Thus, the use of machine learning-based DSS must be coupled with an awareness of the need, from both a legal and humanitarian perspective, to preserve sufficient time and space to allow for human deliberation in decisions on the conduct of hostilities.⁶

5 See, e.g.: U.S. Army Center for Army Lessons Learned, *Civilian Casualty Prevention GTA 90-01-039*, May 2016: https://usacac.army.mil/sites/default/files/publications/GTA%2090-01-039_Civilian_Casualty_Prevention.pdf, accessed on 15 August 2023; see also R. Stewart and G. Hinds, “Algorithms of war: The use of artificial intelligence in decision making in armed conflict”, ICRC Humanitarian Law and Policy Blog, October 2023: <https://blogs.icrc.org/law-and-policy/2023/10/24/algorithms-of-war-use-of-artificial-intelligence-decision-making-armed-conflict/>.

6 “In order for humans to meaningfully play their role, these 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.” <https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913>.

OBSERVATIONS

The ICRC's view is that caution is warranted as militaries consider integrating machine learning into DSS used for decisions on the use of force, particularly when it comes to prescriptive tools that recommend courses of action or make predictions. The ICRC's overall position is that preserving meaningful human control and judgement in decisions that pose risks to the life and dignity of people affected by armed conflict and other situations of violence is essential to upholding ethical values and ensuring respect for applicable laws, including IHL.⁷

The requirement for humans to make legal determinations does not, of course, mean that they cannot utilize technological aids or computerized analyses to inform their decisions on the use of force, as is common today. As noted above, commanders and planners are obliged to assess all sources of available information and, provided they are appropriately designed and used, DSS may offer benefits in this respect. However, relying solely on any one DSS output in order to make a legal assessment – such as whether an object meets the definition of a military objective at any given time and can be lawfully targeted – is unlikely to satisfy this obligation.

DSS, including machine learning-based DSS, must remain tools that help and support, rather than hinder or displace, human decision-making. Ultimately legal obligations and ethical responsibilities bear on moral agents – human beings – and must not be outsourced to software, however computationally intensive the algorithms may become. To ensure that human decision makers are in a position to, and do in fact, exercise sufficient independent scrutiny of DSS outputs, measures should be implemented in the design and use of the system, including technical standards, such that:

- users of DSS – particularly machine learning-based DSS – are able to understand and challenge their outputs. This requires accounting for the technical capabilities and limitations of the DSS in the circumstances of use, including embedded uncertainties, assumptions and biases that operate to reduce predictability, explainability and understandability
- sufficient time and space are preserved for human judgement and deliberation in decisions on the conduct of hostilities.

Such measures could include designing and using DSS in such a way that they: provide decision makers with the type, quality and quantity of information that in practice facilitate and improve situational understanding; allow for output to be cross-checked against another source of information; enable users to adequately account for the technical and human-machine interaction challenges inherent to DSS, particularly those that are machine-learning enabled; and afford users the time required for deliberation and the practical possibility to exercise human judgement in a contextual manner.

Finding ways to achieve this both through technical features and operating procedures that take into account enduring human-machine interaction challenges will be key to ensuring that DSS can deliver on military claims of supporting improved IHL compliance, even while they might also produce a military advantage.

The ICRC hopes that the enclosed report will contribute to international discussions on the use of DSS in conflict, and that it contains useful insights and practical considerations for states and other actors developing and using these systems.

⁷ “It is essential to preserve human control and judgement in applications of AI and machine learning for tasks and in decisions that may have serious consequences for people's lives, especially where these tasks and decisions pose risks to life, and where they are governed by specific rules of international humanitarian law. AI and machine learning systems remain tools that must be used to serve human actors, and augment human decision-makers, not replace them.” <https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913>.

DECISIONS, DECISIONS, DECISIONS: COMPUTATION AND ARTIFICIAL INTELLIGENCE IN MILITARY DECISION-MAKING

**Prepared by Arthur Holland Michel for the International Committee of the Red Cross (ICRC).
This report does not necessarily represent the view of the ICRC.**

CONTENTS

| | |
|--|-----------|
| Abbreviations | 12 |
| 1. Introduction | 13 |
| 1.1 Key Takeaways of this Report | 16 |
| 2. What are Decision Support Systems? | 17 |
| 2.1 Perceived Military Benefits of Decision Support Systems | 18 |
| 2.2 Potential Risks of Decision Support Systems | 21 |
| 3. Anticipated Emerging Decision Support Systems Capabilities and Use | 23 |
| 3.1 Advances in Information Processing and Planning | 23 |
| 3.2 Speed and Scale | 26 |
| 3.3 Adaptability | 27 |
| 3.4 Expanding Use of Machine Learning-Based Decision Support Systems | 28 |
| 4. Uncertainties, Assumptions and Biases in Decision Support Systems | 31 |
| 4.1 Uncertainties | 32 |
| 4.2 Assumptions | 36 |
| 4.3 Biases | 40 |
| 4.4 The Varying Relevance of Uncertainties, Assumptions and Biases | 43 |
| 4.5 Human Limitations | 45 |
| 5. Implications of Complex Decision Support Systems for Decision-Making in the Use of Force | 51 |
| 5.1 The Shrinking Space for Human Intervention | 51 |
| 5.2 Accountability in Decision-Making | 53 |
| 5.3 Unpredictability, Errors and Cyber Vulnerabilities | 54 |
| Annex: Roles of Decision Support Systems in the Use of Force | 55 |

ABBREVIATIONS

| | |
|--------------|--------------------------------|
| AI | Artificial Intelligence |
| CDE | Collateral Damage Estimation |
| COA | Course of Action |
| DSS | Decision Support System |
| IHL | International humanitarian law |
| VBIED | vehicle-borne explosive device |

SECTION 1

INTRODUCTION

In warfare, the process leading to the use of force is punctuated by many critical human decisions.¹ None of these decisions is easy.² Regardless of whether the process stretches across minutes or weeks, those making these decisions must take into account a constellation of complex factors. These include evolving intelligence assessments and uncertainty³ about the environment and the people in that environment, be they the civilian population, the adversary or their own forces; the over-arching strategic goals with 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 these many variables into account, decision makers must seek to maximize the probability of achieving the objective with the lowest possible risk of adverse or unintended outcomes, while also – crucially – complying with all relevant international humanitarian law (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,⁴ they directly and often significantly influence the decisions of human decision makers.⁵ As a result of advances in areas like computing, artificial intelligence (AI) – especially machine learning – data collection and communications, their capabilities will grow significantly in the years ahead, as will their influence on military decision-making. This report is intended to illuminate the functions of military DSS – in particular, increasingly machine learning-based DSS – and highlight their implications and some key limitations that could be relevant for the application of the law to human decisions in the process leading to the use of force.

-
- 1 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. ICRC, *Decision-Making Process in Military Combat Operations*, ICRC, 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?”. G. Desclaux and B. 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.
 - 2 A. Tolk and D. Kunde, “Decision Support Systems – Technical Prerequisites and Military Requirements,” 2000 Command and Control Research and Technology Symposium, June 2000, Monterey, CA, United States.
 - 3 H. Atoyan, J.-M. Robert and J.-R. Duquet, “Uncertainties in complex dynamic environments,” *Journal d’Interaction Personne-Système*, Vol. 2, No. 1, Art. 5, January 2011: <http://www.indiandefencereview.com/spotlights/uncertainty-and-risk-in-military-decision-making/>.
 - 4 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. Decision Support Systems, which always support a human decision, are distinct from Decision Systems that make automated decisions. M. Bohanec, *What is Decision Support?*, Jožef Stefan Institute, Ljubljana, 2001, p. 2.
 - 5 M. 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, Art. 6, 2018, p. 23.

DSS are thought 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.⁶ Compared to non-computerized methods for supporting a decision, DSS are regarded as being faster,⁷ more comprehensive, more efficient, more consistent⁸ and less prone to errors.⁹ Therefore, the ICRC has previously noted that such tools “may enable better decisions by humans in conducting hostilities in compliance with international humanitarian law and minimizing risks for civilians by facilitating quicker and more widespread collection and analysis of available information.”¹⁰ 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 that the intentions of the humans operating the systems are aligned with those norms.¹¹

However, the ICRC has also observed that the “use and misuse” of DSS “could lead to increased risks for civilian populations.”¹² 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 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 overreliance on computerized analyses and predictions might “facilitate worse decisions or violations of international humanitarian law and,” likewise, “exacerbate risks for civilians.”¹³

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, in particular, developments in machine learning, these systems are becoming more complex and will be used more widely to carry out 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 DSS, including those incorporating machine learning, could reduce and hinder the application of human judgement in decisions on the use of force, and thus shrink the space for human intervention in the processes¹⁴ of conflict.

-
- 6 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, Cologne, 2019, p. 11; 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.
 - 7 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. R. Rasch, A. Kott and K.D. Forbus, “Incorporating AI into military decision making: an experiment”, *IEEE Intelligent Systems*, Vol. 18, Issue 4, July–August 2003.
 - 8 M. 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, Art. 6, 2018, p. 24; W.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.
 - 9 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; C. Godé and J-F. 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, No. 2, 2013.
 - 10 ICRC, “Artificial intelligence and machine learning in armed conflict: A human-centred approach”, *IRRC*, No. 102 (913), Digital technologies and war, 2020, pp. 463–479.
 - 11 Interview with Margarita Konaev, October 2021.
 - 12 ICRC, “Artificial intelligence and machine learning in armed conflict: A human-centred approach,” *IRRC*, No. 102 (913), Digital technologies and war, 2020, pp. 463–479.
 - 13 *Ibid.*
 - 14 That is, the degree to which *human* agents can be held accountable for harms.

Modern machine learning, which has yet to be employed widely in critical DSS roles directly implicated in the process leading to the use of force,¹⁵ is likely to pose additional challenges in this regard, especially in terms of bias, predictability and understandability. These challenges could be of 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 constrained.

DSS are also directly relevant to the ongoing debate regarding autonomous weapon systems. Though DSS stand apart from weapons that “select and apply force to targets without human intervention,”¹⁶ many of the tools that are currently used or will be used in the future to support human decisions in the use of force (such as automatic target recognition tools, planning and optimization tools) could play a critical function in autonomous weapon systems that execute those same “decisions” autonomously. As such, the well-established limitations and risks of DSS are likely to prefigure the limitations and risks of future autonomous weapons systems. Furthermore, any requirement for human control and judgement in specific attacks, and supervision of the critical targeting functions of an autonomous weapon system, would directly implicate many of the same challenges that attend human-DSS interaction today, as well as the novel challenges that future technological developments will bring to bear on these interactions in the years ahead.

In contrast to autonomous weapons, the use of DSS in conflict has received relatively scant attention. This report is intended to serve as a foundation for dialogue on the issues that DSS pose and what to do about them. First, it describes the roles and perceived benefits of modern decision-support tools, as well as a discussion of their risks (Section 2). It then provides a general forecast of how their capabilities and use are likely to expand as a result of advances in computing, sensing, communications and AI (Section 3). In Section 4, it discusses the challenges of assuring that human users make the right decisions when interacting with DSS systems whose outputs are inevitably marked by context-dependent uncertainties, assumptions and biases, and indicates how these challenges grow significantly with the advent of machine learning-based DSS (Section 4). Finally, it discusses how the growing use of these complex DSS is likely to hinder the faithful application of the law to military decisions on the use of force and holding those involved accountable. (Section 5).

¹⁵ Interview with Peter Svenmarck, November 2021; P. 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.

¹⁶ ICRC, *ICRC Position on Autonomous Weapon Systems*, ICRC, Geneva, 2021.

1.1 KEY TAKEAWAYS OF THIS REPORT

- DSS are computerized tools that are designed to aid humans in making complex decisions by presenting information that is relevant for the decision or by proposing options for the decision maker to choose from in order to achieve a goal. DSS can play a critical role in decisions leading to the use of force.
- DSS are perceived to be beneficial for improving the quality, increasing the speed and bolstering the consistency of human decisions.
- In the years ahead, technological advances, especially in machine learning, are expected to both increase the performance of DSS and expand their use in conflict. These advances are also expected to “automate” more aspects of decision-making by reducing the layers of human reasoning and judgement that would have been required in decisions based on simpler DSS outputs.
- Given that DSS do not “make” decisions, human decision makers must possess the capacity to ensure that the use of DSS does not result in decisions that cause unintended or unlawful harm. This capacity relies, in turn, on the human user’s ability to gauge the *uncertainties*, *assumptions* and *biases* that are embedded in DSS outputs in relation to the unique context of the decision it supports.
- There are inherent obstacles to the capacity of DSS users to grasp and account for uncertainties, assumptions and biases when making decisions. These obstacles grow as DSS become more complex and as they are used for a wider variety of less mathematically definable tasks. In particular, the use of machine learning in DSS significantly expands potential uncertainties, assumptions and biases and expands the challenges of accounting for these factors in human decisions based on DSS outputs.
- Therefore, the *expanding use* of more *complex* DSS, including those incorporating machine learning, in decisions on the use of force is likely to *reduce* and *hinder* the application of human judgement – and thus significantly shrink the space for human intervention in the overall process.
- Increasingly complex DSS raise a range of additional challenges for the application of the law to decisions on the use of force, including issues related to understandability and predictability, errors and cyber vulnerabilities. These issues become more serious as the use of these DSS expands across the processes leading to the use of force.

SECTION 2

WHAT ARE DECISION SUPPORT SYSTEMS?

DSS are computerized tools that are designed to aid humans in making complex decisions by presenting information that is relevant for the decision or proposing options for the decision maker to choose from in order to achieve a goal.

DSS are widely used in a range of fields, from medicine to business administration and logistics.¹⁷ The technology is even becoming common in everyday life. Product recommendation engines for online shopping are a type of DSS that speeds up the search for products that (in theory) most closely match one's needs. Travel websites help users decide where to dine by collating information on restaurants' locations, pricing and user reviews. Navigation software proposes a route for how to reach a destination and provides information that is relevant to the journey (such as traffic jams or the presence of roadworks and speed cameras).¹⁸ Ride-hailing apps have supplanted human dispatchers who previously had to make decisions about what car to send each customer.¹⁹ All of these tasks could, in theory, be carried out "manually," but doing so would require one to process a large amount of information and solve mathematically complex, or at least very tedious, problems.

In military conflict, DSS serve similar functions across a wide range of tasks at many different levels of the chain of command. Such capabilities may be employed in service of decisions throughout the process leading to the use of force, both in offense and in defense on every timescale, from pre-selected and pre-planned targeting (sometimes called "deliberate targeting") operations that may span weeks or months to time-constrained targeting operations (sometimes called "dynamic targeting") and actions that must be taken in a matter of seconds, such as air defense.

¹⁷ K. McKendrick, *The Application of Artificial Intelligence in Operations Planning*, NATO Science & Technology Organization, Brussels, 2017, STO-MP-SAS-OCS-ORA-2017, p. 4.

¹⁸ Military officials have likened their envisioned future Command and Control networks to operate similarly to the crowdsourced navigation app Wayze. See: T. Hitchens, "MDO Exclusive: Air Force Targets Primary Role in Joint C2", *Breaking Defense*, 21 January 2020: <https://breakingdefense.com/2020/01/mdo-exclusive-air-force-targets-primary-role-in-joint-c2/>.

¹⁹ V.N. Gadepally *et al.*, "Recommender Systems for the Department of Defense and Intelligence Community," *Lincoln Laboratory Journal*, Vol. 22, No. 1, 2016.

Deterministic vs. Non-deterministic Decision Support Systems

Many traditional DSS are based on “deterministic” (rule-based) computer models and algorithms that will always produce the same output for the same input. For example, a DSS tool that calculates the range of an aircraft based on its fuel levels and speed uses a simple formula that will always function the same way when given the same inputs. While such systems are predictable and understandable for the user, they are often described as being limited in their capacity to process complex problems, account for unobserved variables, compute large numbers of conditions and parameters,²⁰ solve “unstructured” problems and account for dynamics that are impossible to ‘script’ (i.e. code with a specific rule) into the software.²¹

By contrast, machine learning-based DSS are based on “non-deterministic” models also known as “probabilistic” or “stochastic” models, which the computer develops on the basis of training datasets with examples of desired outputs for given inputs. Such systems do not need to be coded with constrained and abstracted rules that may fail to fully capture the complexity of the challenges they are intended to solve.²² As a result, their performance on such tasks might be better. However, because these models include elements of randomness to account for variables in the environment, they may generate different outputs for the same or similar inputs. Therefore, it may be unclear why the system has produced a given output, and it may be harder to predict exactly how they will function in any given instance of use.

2.1 PERCEIVED MILITARY BENEFITS OF DECISION SUPPORT SYSTEMS

In the process leading to the use of force, there are numerous functions and steps in which DSS might play a role. For a description of these roles, see the Appendix. In these roles, DSS are perceived to be beneficial for improving the quality, increasing the speed and bolstering the consistency of human decisions.

For example, a commander wishing to destroy a target would typically need to locate and positively identify that target and seek information about its surroundings that would be relevant to the attack, including the consequences for civilians and civilian objects that may be affected, as well as adversary forces who might defend or retaliate against the operation. In this process of

20 For example, it has been observed that rule-based DSS from the 1980s become un-tenable and “un-maintainable” if too many rules were added, while game-based simulation and modelling systems “gradually became hopelessly convoluted” as more dynamics of the environment were added. P. 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. 2.

21 O. Leifler, “Affordances and Constraints of Intelligent Decision Support for Military Command and Control— Three Case Studies of Support Systems”, Linköping Studies in Science and Technology Dissertation No. 1381, 2011. For example, traditional collateral damage estimation tools “cannot always account for the dynamics of the operational environment,” according to: U.S. Department of Defense, *No Strike and the Collateral Damage Estimation Methodology*, Chairman of the Joint Chiefs of Staff Instruction CJCSI 3160.01, 13 February 2009. Such systems may not, for example, be able to account for moving civilian individuals or vehicles that are passing near the target at the time of a strike. S. Muhammedally, “Minimizing civilian harm in populated areas”, *IRRC*, No. 98 (1), 2016, p. 244; interview with Lawrence Lewis, September 2021; anonymous interview with an NGO employee, September 2021; B.S. Lambeth, *Air Power Against Terror: America’s Conduct of Operation Enduring Freedom*, RAND Corporation, Santa Monica, 2006, pp. 320–321.

22 P. Svenmarck et al., *Possibilities and Challenges for Artificial Intelligence in Military Applications*, NATO Science & Technology Organization, STO-MP-IST-160-S1-5P, Brussels, 2018, p. 2; Lt. Col. G., Major (Res.) G. and Major (Res.) L., “From Traffic Analysis to Artificial Intelligence”, *Dado Center Journal*, 2 March 2021: <https://www.idf.il/en/minisites/dado-center/research/from-traffic-analysis-to-artificial-intelligence/>.

information retrieval and discovery, DSS are regarded as being helpful for helping make decision makers aware of all the information that is relevant to their decision without having to manually review every bit of information that may be relevant.²³ DSS may also offer the benefit of presenting that information in a way that is more intuitive and easier to “understand” than raw data.²⁴

Similarly, when developing a “course of action” (COA) to achieve a goal, decision makers must optimize a large number of variables and constraints related to available resources for executing the plan, environmental features, adversary forces and the civilian population. Furthermore, they must ensure that any action remains in keeping with all top-level requirements derived from national and international laws, rules of engagement and directives.²⁵ In such planning and optimization tasks, computerized tools may be perceived to be the only means of running all of a potential plan’s numerous variations and accounting for *all* known variables in time to enable decision makers to consider their options before taking a decision and actually carry out the mission,²⁶ particularly in cases where the time available to make a decision is extremely short.²⁷

During the execution of an attack, decision makers will typically continue to seek information on relevant developments (such as the target moving to a new location, civilians entering the area or the emergence of risks to their own forces) that may, in some cases, require the attack to be suspended or cancelled. This process requires both information monitoring as well as planning. Following the attack, analysts will seek information to determine whether the objective was achieved and, if needed, develop a new plan for how to proceed (for example, by re-attacking the target or launching a post-strike investigation).

²³ S.C. Gordon, “Decision Support Tools for Warfighters”, 2000 Command and Control Research and Technology Symposium, Monterrey.

²⁴ For example, a map with the locations of all known enemy positions is more intuitive than a list of those positions’ coordinates.

²⁵ For a vignette describing how a variety of decision support tools could be used in the use of force in a deliberate targeting operation, see: N. Gizzi, J. McDonnell and A. Rice, “The State of the Art and the State of the Practice Dynamic Decision Support for Time Critical Targeting”, 2006 Command and Control Research and Technology Symposium, San Diego.

²⁶ Interview with Herman le Roux, November 2021; P-I. Evensen and D.H. Bentsen, *Simulation of land force operations – a survey of methods and tools*, Norwegian Defence Research Establishment, Oslo, 15 February 2016, p. 14.

²⁷ For example, in air defense roles, some form of automatic detection of possible target is deemed necessary by militaries because the limited crew available for such a task would not be capable of scanning the whole sky manually: Д.В. Галкин, П.А. Коляндра and А.В. Степанов, “Состояние и перспективы использования искусственного интеллекта в военном деле”, *Военная Мысль*, No. 1, p. 115; interview with Milind Kulshreshtha, September 2021.

EXAMPLES OF DECISION SUPPORT SYSTEMS

A **course of action evaluation tool** that provides information on the options available for achieving a goal.



A **resource optimization tool** that indicates the “optimal” positions for weapons or the available weapons that are “optimal” to carry out a given attack.



A **recommender system** that prioritizes information that may be of higher relevance from a larger pool of data.

A **data analytics tool** that uncovers statistical patterns in data to identify information of military relevance.



A **geographic information system** that enables operators to compare areas.



A **mapping system** that indicates the location and characteristics of geographic features.

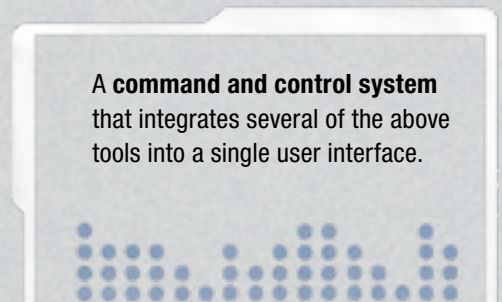


A **route planning tool** that indicates the different routes that a force can take to move from point A to B.



A **biometric identification system** that matches the characteristics of an individual to an identity.

A **command and control system** that integrates several of the above tools into a single user interface.



A **defense system** that alerts users to potential incoming missiles, drones or fighter jets.



An **optimization tool** that indicates the “optimal” manner in which to use or allocate resources during an operation.

A **simulation tool** that indicates how an adversary is likely to respond to potential courses of action.



A **predictive tool** that indicates how probable it is that a given event will occur in the future.



A **collateral damage estimation tool** that provides an estimate of the likely effect of deploying a given weapon against a particular target.

2.2 POTENTIAL RISKS OF DECISION SUPPORT SYSTEMS

DSS can be used to improve the quality of decision-making in the use of force, and therefore support the application of rules of engagement and facilitate compliance with the requirements of international law in attacks.²⁸ However, it is also important to have a realistic assessment of the limitations of DSS.

Like all computerized systems, DSS can experience failures, as can the people, organizations, policies and processes that employ them. Human-machine interaction challenges also pose limitations. Systems that display incorrect information or that fail to display relevant information can cause decision makers to draw incorrect conclusions and make decisions that cause unintended or unlawful harm.²⁹ Planning tools that generate suboptimal solutions in relation to key constraints may spur decision makers to make decisions that fail to align with requirements, such as IHL rules and rules of engagement pertaining to the protection of civilians.

DSS can also be used inappropriately even when they are operating exactly as designed. For example, if a human uses the output as the sole basis for deciding to launch an attack, even though that system is only capable of providing incomplete information or a solution that cannot account for relevant constraints in that context.

When these tools are brought to bear on decisions on the use of force, they can expand the exposure of civilian populations to risk. These risks can be substantial, whether or not the system is used in close proximity (temporally, materially or otherwise) to the decision on the use of force.

For example, an intelligence synthesis tool may contribute to the failure of analysts to see pieces of information that will be highly relevant at a later stage of the process leading to the use of force.³⁰ A planning tool might result in a potentially inappropriate weapon system being deployed (say, a weapon with high fragmentation deployed to a target that is in proximity to civilians), even though it had no direct role in the decision maker's ultimate decision to launch the attack. A mapping system might fail to indicate the correct location of a relevant object, such as a civilian structure, which might throw off all subsequent analysis and planning related to that area.

In rapid, tightly linked "kill chain" processes leading to the use of force, seemingly innocuous failures can elevate the risk of what is known as a 'cascading error' that perpetuates through all subsequent functions. There is also speculation that the use of complex decision support at the national command level may contribute to strategic miscalculations³¹ or runaway escalation between states, and even increase the risk of use of nuclear weapons.³²

²⁸ "Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task"; A. Deeks, N. Lubell and D. Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States", *Journal of National Security Law & Policy*, Vol. 10:1, p. 10; M. 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, Art 6, 2018.

²⁹ For example, if an automatic target recognition tool misidentifies a civilian aircraft as a military aircraft – or if an information synthesis or retrieval tool fails to display intelligence regarding civilian structures in proximity to a proposed target.

³⁰ M. 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, Art 6, 2018, p. 80.

³¹ M. Horowitz and P. Scharre, *AI and International Stability: Risks and Confidence-Building Measures*, Center for a New American Security, Washington, 12 January 2021.

³² S.A. Sial, "Military applications of artificial intelligence in Pakistan and the impact on strategic stability in South Asia", in P. Topychkanov (ed.), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Vol. III, Stockholm International Peace Research Institute, April 2020; O. Daniels, "Speeding Up the OODA Loop with AI A Helpful or Limiting Framework?", Joint Air & Space Power Conference 2021; B.J. Schachter, *Automatic Target Recognition*, Vol. 4, SPIE, Bellingham, 2020, p. 298.

That being said, **a decision is a human act that can only be carried out through human judgement. DSS do not make decisions.** Even if the human does exactly what a DSS proposes that they do, it is still a decision. A human should be able to make the “right” decision even if the output of the system is erroneous or inappropriate.

Conversely, humans can make harmful decisions even when supported by DSS that are operating normally and without error. To ensure that DSS do not result in unintended harm, it is critical that human decision makers always maintain their capacity to make appropriate, contextual decisions on the basis of any DSS output. This is discussed further in Section 4.

SECTION 3

ANTICIPATED EMERGING DECISION SUPPORT SYSTEMS CAPABILITIES AND USE

The coming years will see significant growth in the capabilities of DSS as a result of technology developments in sensing, communications, computing and AI, especially machine learning-based algorithms³³ and foundation models. These technologies are seen as a means to enhance the aforementioned benefits deemed to be expected from DSS by expanding the volume and diversity of information that DSS can process, increasing the complexity of tasks that these tools can carry out and increasing the speed and flexibility of DSS-enabled human decision-making.³⁴

These advances are also expected to “automate” more aspects of decision-making by reducing the layers of human reasoning and judgement that would have been required in decisions based on simpler DSS outputs. However, as noted in Section 5, many of these advances are also likely to challenge the application of human judgement to the outputs of DSS.

3.1 ADVANCES IN INFORMATION PROCESSING AND PLANNING

The developments mentioned above are expected to enable the extraction of more information from available sources so that it can be more readily retrieved, fused and analysed.³⁵ This could enable DSS to make a greater volume and variety of information available to decision makers. For example, a mapping tool or planning tool might incorporate recent satellite imagery, com-

³³ The research community has spent decades working to leverage forms of “artificial intelligence” for DSS. For a detailed literature review that charts the development of AI for DSS, in particular for optimization problems, between 1975 and 2015, see A. Naseem *et al.*, “Decision support system for optimum decision making process in threat evaluation and weapon assignment: Current status, challenges and future directions”, *Annual Reviews in Control*, 43, 2017, pp. 169–187. For an overview of types of “AI” that are commonly applied for DSS, see: W. Wang *et al.*, “Investigation on Works and Military Applications of Artificial Intelligence”, *IEEE Access*, Vol. 8, 2020, pp. 131,614–131,625; see also: W. Wiseman, *Deep Learning for Human Decision Support*, Defence Research and Development Canada, Ottawa, 20 January 2017, and M. Bistrion and Z. Piotrowski, “Artificial Intelligence Applications in Military Systems and Their Influence on Sense of Security of Citizens”, *Electronics*, 2021, 10, p. 871.

³⁴ For example, A. Рамм and A. Козаченко, “Командир на автопилоте: управлять армиями поможет компьютер”, *Iz.ru*, 5 June 2019, <https://iz.ru/884970/aleksei-ramm-aleksei-kozachenko/komandir-na-avtopilote-upravliat-armiiami-pomozhet-kompiuter>.

³⁵ For example, a facial recognition system, could “extract” from a video clip a list of (suspected) individuals who resemble those individuals in the footage. A “re-identification” computer vision algorithm could track unknown individuals as they appear in disparate camera feeds. Speech recognition systems could transcribe radio chatter into searchable text. A system that can tag intelligence documents or memoranda according to their type and content could potentially aid in faster and more expansive intelligence synthesis, prioritization, and visualization. M. Chen, Z. Wang and F. Zheng, “Benchmarks for Corruption Invariant Person Re-identification”: <https://arxiv.org/abs/2111.00880>; “BABEL”, University of Cambridge Department of Engineering: <http://mi.eng.cam.ac.uk/~mjfg/BABEL/index.html>; J. Schubert *et al.*, “Artificial Intelligence for Decision Support in Command and Control Systems”, 23rd International Command and Control Research & Technology Symposium, Playa Vista, 2017, pp. 5–6.

mercially available location data from smartphones,³⁶ biometric data or information about the presence and activities of people based on social media analytics.³⁷ Meanwhile, biometric identification or verification techniques are being increasingly used to estimate a human's identity on the basis of physical, physiological or behavioural features, such as their fingerprint, their DNA, their facial geometry or their gait.³⁸

Meanwhile, more advanced analytics is expected to elevate the function of analytical tools from simply measuring or calculating mathematically defined characteristics of objects or phenomena³⁹ to "predicting" fuzzier latent characteristics such as behaviour,⁴⁰ intent, relationship to other entities,⁴¹ current and future state⁴² and other predictions.⁴³ Such "predictive" systems draw on hard data to "infer" facts that are not directly observed.

For example, a system with such a capability might be used to predict the destination of a convoy by measuring its size, direction and speed of travel.⁴⁴ A system might attempt to predict the likelihood that a person is an enemy fighter not only based on their wearing a uniform or carrying a weapon, but also based on their social connections to individuals who are known or suspected of being enemy fighters.

-
- 36 M. Clark, "US Defense Intelligence Agency admits to buying citizens' location data", The Verge, 22 January 2021: <https://www.theverge.com/2021/1/22/22244848/us-intelligence-memo-admits-buying-smartphone-location-data>.
- 37 PTE. E., "The Tactical Application of Open Source Intelligence (OSINT)", The Cove, 27 October 2020: <https://cove.army.gov.au/article/tactical-application-open-source-intelligence-osint>.
- 38 Interview with Svetlana Yanushkevich, November 2021; "Biometrics", National Institute of Standards and Technology: <https://csrc.nist.gov/glossary/term/biometrics>.
- 39 Say, an object's shape, location, speed, color, and anomalies in structured datasets.
- 40 See, for example: Lt. Col. G., Major (Res.) G. and Major (Res.) L., "From Traffic Analysis to Artificial Intelligence", *Dado Center Journal*, 2 March 2021: <https://www.idf.il/en/minisites/dado-center/research/from-traffic-analysis-to-artificial-intelligence/>; G. Sreenu and M.A. Saleem Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis", *Journal of Big Data*, Vol. 6, Art. 48, 2019.
- 41 Rather than identifying an individual generically as a "combatant," the hope is that such a system might, for example, identify their level of seniority. A. Bergeron Guyard, *Self-improving inference system to support the intelligence preparation of the battlefield: Requirements, state of the art, and prototypes*, Defence Research and Development Canada, Scientific Report DRDC-RDDC-2014-R136, Ottawa, December 2014, pp. 17–19.
- 42 B. Cook, "The Future of Artificial Intelligence in ISR Operations", *Air & Space Power Journal*, Vol. 35, Special Issue – Perspectives on JADO, Summer 2021; "The Human-Machine Team: How to Create Synergy Between Human & Artificial Intelligence That Will Revolutionize Our World".
- 43 K. McKendrick, *The Application of Artificial Intelligence in Operations Planning*, NATO Science & Technology Organization, Brussels, 2017, STO-MP-SAS-OCS-ORA-2017, p. 7. For example, according to one report, while Korea's existing command and control tool, ATCIS, can aid decision making by displaying and consolidating "the actual facts of the battlefield" for human decision makers, "the principal decision making, such as 'the possibility of hostile provocation' and 'the most effective strike method' depends on the intuition and experience of the commanders and staff officers": D. Yoo, S. No and M. Ra, "A Practical Military Ontology Construction for the Intelligent Army Tactical Command Information System", *International Journal of Computers Communications & Control*, 9 (1), pp. 93–100; Margarita Konaev (interviewed October 2021) noted that a principal difference between legacy DSS and Machine Learning-enabled systems will be the newer system's capacity for predictive (rather than merely descriptive) analytics. ⁵⁰ S.C. Gordon, "Decision Support Tools for Warfighters," 2000 Command and Control Research and Technology Symposium, Monterrey, p. 8; J. Barker et al., "Information Fusion Based Decision Support via Hidden Markov Models and Time Series Anomaly Detection", 12th International Conference on Information Fusion, Seattle, 6–9 July 2009; see also: "Striking Smarter and Faster," Singapore Defence Science & Technology Agency: <https://web.archive.org/web/20200810083701/https://www.dsta.gov.sg/programme-centres/information-pc/striking-smarter-and-faster>; F. Wang, "Technology Framework of the Intelligent Command and Control System", *Materials Science and Engineering*, 677, 2019.
- 44 Data analysis and fusion tools such as the AIRBUS Fortion ABI-based ISR systems are designed to be used to "identify" targets on the basis of their activities rather than solely their physical characteristics.

So-called “soft biometrics” such as “gait recognition” or “emotion recognition”⁴⁵ that predict abstract qualities such as emotional state or intent are similarly a growing area of research, though there is broad scientific consensus that such techniques can be inaccurate and pose a high risk of perpetuating harmful toxic biases.⁴⁶ Military practitioners hope that these types of tools will lead to more granular threat detection and object/event identification based on incomplete information or intangible qualities,⁴⁷ which they expect to result in significant improvements in “situational awareness.”⁴⁸

In the sphere of systems that propose choices for decision makers, AI and related technologies are predicted by some to deliver significant gains by accounting for a wider number of variables and options.⁴⁹ Machine learning-based systems can, in theory, account for and optimize for latent variables that are not directly “coded” into the system by their designers.

There is even speculation that some DSS could outstrip human capacity for strategic and tactical planning.⁵⁰ For example, in early demonstrations involving complex strategic board games like Go and computer games like StarCraft, reinforcement learning-based systems have been demonstrated to be capable of generating solutions for an inconceivable range of scenarios and do so in ways that are (from a human perspective) unintuitive and thus difficult for an adversary to forestall.⁵¹ That being said, it has not been demonstrated that these same capabilities could be

45 Interview with Svetlana Yanushkevich, November 2021; Erica Wiseman, *Deep Learning for Human Decision Support*, Defence Research and Development Canada, Ottawa, 20 January 2017, p. 22.

46 Kate Crawford, “Artificial Intelligence is Misreading Human Emotion”, *The Atlantic*, 27 April 2021: <https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/>.

47 See for example: B. Cook, “The Future of Artificial Intelligence in ISR Operations”, *Air & Space Power Journal*, Vol. 35, Special Issue – Perspectives on JADO, Summer 2021; A. Bergeron Guyard, *Self-improving inference system to support the intelligence preparation of the battlefield: Requirements, state of the art, and prototypes*, Defence Research and Development Canada, Scientific Report DRDC-RDDC-2014-R136, Ottawa, December 2014, pp. 11–12.

48 “Situational awareness” is a widely used term in a range of fields that describes a human decision makers’ perception of their environments and their understanding of the significance of the factors in this environments. M.R. Endsley, “Toward a Theory of Situation Awareness in Dynamic Systems”, *Human Factors*, 37(1), 1995, pp. 32–64. For a critical discussion of the notion of “situational awareness” as applied to military targeting, see: L. Suchman, “Algorithmic warfare and the reinvention of accuracy”, *Critical Studies on Security*, Vol. 8, Issue 2, 2020.

49 袁艺 高冬明 张玉军, “也谈智能化指挥自主决策,” 81.cn, 18 April 2019: https://www.81.cn/jfjbmap/content/2019-04/18/content_231979.htm; M. Walsh et al., *Exploring the Feasibility and Utility of Machine Learning-Assisted Command and Control, Volume I, Findings and Recommendations*, RAND Corporation, Santa Monica, 2021. For an overview of various approaches to applying AI for simulations, see: M. Bistrion and Z. Piotrowski, “Artificial Intelligence Applications in Military Systems and Their Influence on Sense of Security of Citizens”, *Electronics*, 2021, 10, p. 871; E. Wiseman, *Deep Learning for Human Decision Support*, Defence Research and Development Canada, Ottawa, 20 January 2017; A. Пешков, “Шойгу: в МО РФ создана система прогнозирования вооруженных конфликтов”, *Звезда*, 16 December 2019: <https://tvzvezda.ru/news/201912161125-PViRZ.html>, accessed via Google Translate.

50 S. Soleyman and D. Khosla, “Multi-Agent Mission Planning with Reinforcement Learning”, *Proceedings of AAAI Symposium on the 2nd Workshop on Deep Models and Artificial Intelligence for Defense Applications: Potentials, Theories, Practices, Tools, and Risks*, November 2020; M.R. Voke, “Artificial Intelligence for Command and Control of Air Power”, Wright Flyer Paper No. 72, Air University Press, Maxwell Air Force Base, 2019, pp. 15–16; T. Doll et al., *From the Game Map to the Battlefield – Using DeepMind’s Advanced AlphaStar Techniques to Support Military Decision-Makers*, NATO Science & Technology Organization, Brussels, 2021, STO-MP-MSG-184.

51 In the civilian domain, a classic example of this effect was observed in the reinforcement learning-based AlphaGo system, which executed unusual strategies to beat champion human players at the board game Go. For discussion of RL as it applies to mission planning, see: J. Schubert et al., “Artificial Intelligence for Decision Support in Command and Control Systems”, 23rd International Command and Control Research & Technology Symposium, Playa Vista, 2017; P. 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; Z. Xiaohai and C. Xinwen, “Military Intelligent Decision Support System Based on Deep Learning”, *Command, Control and Simulation*, 2018.

translated to a real-world military scenario, which involves far more complexity, variability and chaos than even the most complex virtual world of a computer game.⁵²

3.2 SPEED AND SCALE

Across these functions, the use of machine learning and more powerful computing is also anticipated to make the decision-making cycle faster.⁵³ This would be principally achieved by reducing the volume of information that decision makers would have to scrutinize. For example, instead of displaying every vehicle in an area, a system would display only those vehicles whose “behaviour” matches a pattern that is “relevant” to the decision maker; or instead of reading every intelligence report about a subject, an analyst could use a DSS that synthesizes all available intelligence in a single summary. In planning functions, systems might generate plans automatically without having to be manually ‘scripted’ with information about resources, constraints and objectives.

Furthermore, systems that link various separate DSS functions in a single tool are expected to drastically reduce the time required to go from collecting information to acting on that information because they will not require human intervention for the various DSS at each stage of the process leading to the use of force. By some estimates, a target search, recognition and analysis activity that previously took hours could be reduced to minutes,⁵⁴ and a process that previously took minutes could potentially be reduced to seconds.⁵⁵ This is assuming that the required time for the decision makers’ process of validating the DSS output would not increase due to the increased complexity of the system and its inputs or due to the requirements of rules of engagement and applicable international law. These requirements may, in fact, put constraints on the admissible speed and condensation of human decision-making.⁵⁶

An important effect of the increased speed of DSS-based decisions is that it could enable militaries to make *more decisions*. For example, if earlier manual processes for identifying potential targets for the use of force took days or weeks, that would naturally limit how many targets a military could attack in a conflict. If a DSS reduces that time to hours or minutes, as some claim, it could multiply the number of targets that a military could identify and, by extension, attack.⁵⁷

52 M. Cummings, “The AI That Wasn’t There: Global Order and the (Mis)Perception of Powerful AI”, Policy Roundtable: Artificial Intelligence and International Security, *Texas National Security Review*, 2 June 2020.

53 O. Daniels, “Speeding Up the OODA Loop with AI A Helpful or Limiting Framework?”, Joint Air & Space Power Conference 2021; Д.В. Галкин, П.А. Коляндра and А.В. Степанов, “Состояние и перспективы использования искусственного интеллекта в военном деле”, *Военная Мысль*, No. 1.

54 L. Xiang, “Artificial intelligence and its impact on weaponization and arms control”, in L. Saalman (ed.), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Vol. II, Stockholm International Peace Research Institute, October 2019; R. Rasch, A. Kott and K.D. Forbus, “Incorporating AI into military decision making: an experiment”, *IEEE Intelligent Systems*, Vol. 18, Issue 4, July–August 2003.

55 The Army’s Project Convergence, Congressional Research Service, Washington, 2 June 2022; 袁艺 高冬明 张玉军, “也谈智能化指挥自主决策,” 81.cn, 18 April 2019: https://www.81.cn/jfjbmap/content/2019-04/18/content_231979.htm.

56 ICRC, “Artificial intelligence and machine learning in armed conflict: A human-centred approach”, *IRRC*, No. 102 (913), Digital technologies and war, 2020, pp. 463–479.

57 A DSS tool called Gospel, which is used by the Israel Defence Forces, is claimed to have increased the number of targets identified for attack. According to a former IDF official, “in the past we would produce 50 targets in Gaza per year. Now, this machine [Gospel] produces 100 targets a single day, with 50% of them being attacked.” H. Davies, B. McKernan and D. Sabbagh, “‘The Gospel’: how Israel uses AI to select bombing targets in Gaza,” *The Guardian*, 1 December 2023: <https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets>.

3.3 ADAPTABILITY

Such tools might also be more dynamic. Whereas traditional computerized decision support tools might only be able to generate a single static output for any given designated task and would have to be re-scripted and re-run if the parameters of the task changed, tools capable of real-time data analytics could potentially update their outputs based on new inputs from the environment.

For example, a dynamic collateral damage estimation (CDE) tool would theoretically be able to constantly adjust its estimates based on changes in the observed environment (say, the target moving to a different place or the emergence of new objects in the blast radius).⁵⁸ A planning tool could potentially be engineered to automatically update a proposed COA or simulation in response to new detected developments in the environment.⁵⁹ A threat assessment tool could adjust the threshold for characterizing an object as a threat based on contextual factors (such as the time of day or the proximity of friendly forces). Some have even suggested that such tools could eventually support sophisticated legally mandated human functions such as the obligation to take all feasible precautions in the choice of means and methods of warfare to avoid or at least minimize incidental civilian harm.⁶⁰

Finally, it is anticipated that the tools themselves could be more closely tailored to their specific context. Thanks to ‘active learning’ machine learning systems,⁶¹ it is anticipated that DSS capabilities could be updated during use. For example, some militaries envision enabling users to train systems in real time in response to system failures or suboptimal solutions,⁶² or to load deployed DSS software with updates from the developer in response to new requirements or environmental factors.

⁵⁸ Present-day CDE tools are generally static, and must be manually re-run if new information becomes available that might change the outcome of the estimate. Anonymous interview with an NGO employee, September 2021.

⁵⁹ N. Gizzi, J. McDonnell and A. Rice, “The State of the Art and the State of the Practice Dynamic Decision Support for Time Critical Targeting”, 2006 Command and Control Research and Technology Symposium, San Diego. Interview with Milind Kulshreshtha, September 2021; S.D. Berrier, “Mission Command Intelligence in Multi-Domain Operations”, *Military Intelligence Professional Bulletin*, 1 July 2019; D. O’Connor, *Dynamic Decision Support – A War Winning Edge*, NATO Science & Technology Organization, Brussels, STO-MP-MSG-111, 2013; Interview with Herman le Roux, November 2021; “Striking Smarter and Faster,” Singapore Defence Science & Technology Agency: <https://web.archive.org/web/20200810083701/https://www.dsta.gov.sg/programme-centres/information-pc/striking-smarter-and-faster>; anonymous interview with an NGO employee, September 2021; A. Пешков, “Шойгу: в МО РФ создана система прогнозирования вооруженных конфликтов”, *Звезда*, 16 December 2019: <https://tvzvezda.ru/news/201912161125-PViRZ.html>, accessed via Google Translate.

⁶⁰ A. Deeks, N. Lubell and D. Murray, “Machine Learning, Artificial Intelligence, and the Use of Force by States”, *Journal of National Security Law & Policy*, Vol. 10:1, p. 10.

⁶¹ In other words, systems whose models can continue to be updated and refined after deployment, in contrast to models that are frozen after testing and validation.

⁶² F. Maymir-Ducharme and R. Ernst, “Advanced C4ISR Platforms with Machine Learning Capabilities”, International Training Technology Exhibition & Conference, Stuttgart, 2018; A. Bergeron Guyard, *Self-improving inference system to support the intelligence preparation of the battlefield: Requirements, state of the art, and prototypes*, Defence Research and Development Canada, Scientific Report DRDC-RDDC-2014-R136, Ottawa, December 2014; Paul McLeary, “Pentagon’s Big AI Program, Maven, Already Hunts Data in Middle East, Africa”, *Breaking Defense*, 1 May 2018: <https://breakingdefense.com/2018/05/pentagons-big-ai-program-maven-already-hunts-data-in-middle-east-africa/>.

3.4 EXPANDING USE OF MACHINE LEARNING-BASED DECISION SUPPORT SYSTEMS

These types of anticipated advances are expected to result in a significant expansion in the use of machine learning-based DSS. Given the rapidly increasing volume of information that militaries collect,⁶³ the advent of communication technologies that can make these sources available to a widely distributed set of actors and the growing complexity and speed of conflict itself, DSS are widely seen to be crucial technology for all future forms of warfare, from ‘near-peer’ conflicts⁶⁴ to asymmetric wars.⁶⁵

These advances are also expected to extend the use of DSS to military roles, operations and users that previously did not employ these tools or only did so to a limited degree. Traditionally, complex DSS tools were predominantly available for only higher levels of command⁶⁶ with large teams of analysts and were primarily used in deliberate targeting operations. Now and in the coming years, similar tools are becoming available to decision makers further down the chain of command, as well as actors engaged in ‘fast-turnaround’ attacks.⁶⁷

-
- ⁶³ T. Bao En, “Swimming In Sensors, Drowning In Data— Big Data Analytics For Military Intelligence,” *POINTER, Journal of the Singapore Armed Forces*, Vol. 42, No. 1, 2016, p. 52; F. Maymir-Ducharme and R. Ernst, “Advanced C4ISR Platforms with Machine Learning Capabilities”, International Training Technology Exhibition & Conference, Stuttgart, 2018; interview with Milind Kulshreshtha, September 2021.
- ⁶⁴ In near peer conflicts between technologically advanced belligerents, the volume of relevant information for any decision will be high, operations will span across multiple domains (air, sea, land, space and cyber), the combinatorial complexity of planning will be enormous, and the speed of decision making is anticipated to be a key determinant of success. T. Doll *et al.*, *From the Game Map to the Battlefield – Using DeepMind’s Advanced AlphaStar Techniques to Support Military Decision-Makers*, NATO Science & Technology Organization, Brussels, 2021, STO-MP-MSG-184, p. 14–12; Д.В. Галкин, П.А. Коляндра and А.В. Степанов, “Состояние и перспективы использования искусственного интеллекта в военном деле”, *Военная Мысль*, No. 1, 2021, p. 115; Interview with Milind Kulshreshtha, September 2021; D. Pedersen *et al.*, “Decision Support System Engineering for Time Critical Targeting”, MITRE Technical Paper, Bedford, 1999, p. 2; H. Irandoust and A. Benaskeur, “Human-Autonomy Teaming for Critical Command and Control Functions”, Defence Research and Development Canada, Ottawa, 2020.
- ⁶⁵ In operations such as counterinsurgencies, DSS are considered to be necessary for processing large amounts of information (again, across physical and digital domains) to find adversaries that might not be “visible” by traditional means, to identify and avoid civilians, and to account for complex societal dynamics. S. Reddy, “Artificial Intelligence in Defence”, India AI, 22 August 2019: <https://indiaai.gov.in/article/artificial-intelligence-in-defence>; R.H. Shultz and R.D. Clarke, “Big Data At War: Special Operations Forces, Project Maven, and Twenty-first-century Warfare”, Modern War Institute, 25 August 2020: <https://mwi.usma.edu/big-data-at-war-special-operations-forces-project-maven-and-twenty-first-century-warfare/>; D. Schmorrow *et al.*, “Applied Use of Socio-Cultural Behavior Modeling and Simulation: An Emerging Challenge for C2”, *Proceedings of the 14th International Command and Control Research and Technology Symposium*, June 2009, Washington.
- ⁶⁶ S.J. Banks, “Lifting Off of the Digital Plateau With Military Decision Support Systems”, Master’s Thesis, School of Advanced Military Studies, United States Army Command and General Staff College, 2013.
- ⁶⁷ For example, complex planning support or simulation capabilities could potentially be deployed in response to “targets of opportunity,” enabling operators to run through a multitude of possible courses of action and to engage in rapid analysis and damage estimation and optimization, within the small window of time that may be available to strike that target (or to withhold from striking it, as the case may be). Interview with Milind Kulshreshtha, September 2021; interview with Herman le Roux, November 2021; A. Goldfarb and J. Lindsay, “Artificial Intelligence in War: Human Judgment as an Organizational Strength and a Strategic Liability”, Brookings Institute, Washington, November 2020; M. Walsh *et al.*, *Exploring the Feasibility and Utility of Machine Learning-Assisted Command and Control*, Volume 1, *Findings and Recommendations*, RAND Corporation, Santa Monica, 2021; P. Tucker, “The US Army Wants to Reinvent Tank Warfare with AI”, Defense One, 18 October 2019: <https://www.defenseone.com/technology/2019/10/us-army-wants-reinvent-tank-warfare-ai/160720/>; А. Рамм and А. Козаченко, “Командир на автопилоте: управлять армиями поможет компьютер”, *Iz.ru*, 5 June 2019: <https://iz.ru/884970/aleksei-ramm-aleksei-kozachenko/komandir-na-avtopilote-upravliat-armiimi-pomozhet-kompiuter>; see also Singapore’s “Smart” Command Post (CP) and the U.S. Defense Advanced Research Projects Agency, Adapting Cross-domain Kill-webs (ACK) program.

These technological developments coupled with the miniaturization of enabling technologies are even expected to bring DSS to actors “at the edge”, who are directly carrying out the use of force. DSS tools embedded in devices such as tablets, smartphones, weapon scopes or augmented reality/heads-up displays⁶⁸ and remote sensors, are expected to be capable of executing a range of DSS tasks, from target detection and tracking to planning.⁶⁹

Advances in communications and miniaturized sensor processing are also expected to facilitate the use of uncrewed systems and sensor networks for remote surveillance. Instead of requiring constant monitoring, such systems would employ, for example, target detection techniques so as to only alert human operators when they encounter something of interest.⁷⁰

Each individual decision maker is also anticipated to have more DSS functionalities at their disposal than before. Individual DSS could consolidate distinct functions in a single output for the decision maker.⁷¹ Mapping tools, for example, could be overlaid with a wider range of available sources of data,⁷² along with descriptive DSS that facilitate their analysis and prescriptive tools that propose decisions on the basis of these data.⁷³ Foundation models might enable a human to coordinate all of these functions through a single chatbot “assistant”.⁷⁴

-
- 68 “Tomorrow’s soldiers will have their reality augmented”, *The Economist*, 22 September 2021: <https://www.economist.com/science-and-technology/tomorrows-soldiers-will-have-their-reality-augmented/21804963>; M. Chmielewski, K. Sapiejewski and M. Sobolewski, “Application of Augmented Reality, Mobile Devices, and Sensors for a Combat Entity Quantitative Assessment Supporting Decisions and Situational Awareness Development”, *Applied Science* 9, 4577, 2019. The integration of DSS into equipment for soldiers is a major component of modernization programs including Nett Warrior (US), IdZ (Germany), FIST (UK), Félin (France), Land 125 (Australia), MARKUS (Sweden), Soldato Futuro (Italy), IMESS (Switzerland), Projekt TYTAN (Poland), FINSAS (India) and ACMS (Singapore), Ratnik (Russia), Advanced Combat Man System (ACMS) (Singapore), SARV (Iran).
- 69 Aircraft cockpits have long incorporated a multitude of computerized decision support tools for tasks such as target detection, GIS, and weaponizing; advances in AI are anticipated to deliver even more extensive DSS capabilities to pilots and onboard operators of vehicles on the ground and in the sea, especially as they are fed with more data from external sources, other systems in the battlespace, or uninhabited systems with which they are teaming. A.P. Chowdhury, “How are Fighter jets embracing Artificial Intelligence?”, *Analytics India Magazine*, 25 July 2017: <https://analyticsindiamag.com/fighter-jets-embracing-artificial-intelligence/>.
- 70 These tools can also reduce the quantity of data that militaries must channel across their limited (and vulnerable) communications links; a target detection system will only transfer those portions of a sensor feed that are relevant back to the human commander rather than sending back the entire data load. Interview with Margarita Konaev, October 2021; anonymous interview with government employee, October 2021; interview with Milind Kulshreshtha, September 2021; R.H. Shultz and R.D. Clarke, “Big Data At War: Special Operations Forces, Project Maven, and Twenty-first-century Warfare”, *Modern War Institute*, 25 August 2020: <https://mwi.usma.edu/big-data-at-war-special-operations-forces-project-maven-and-twenty-first-century-warfare/>.
- 71 For example, a single system might detect a target, estimate the role that target lays within the broader target system, simulate a range of outcomes for a variety of courses of action to use force against that target, and select an optimum plan by which to kill that target (or withhold from targeting it); the human operating this system will thus receive a single output recommending a specific course of action.
- 72 Д.В. Галкин, П.А. Коляндра and А.В. Степанов, “Состояние и перспективы использования искусственного интеллекта в военном деле”, *Военная Мысль*, No. 1, p. 115.
- 73 This is a principle, for example, behind the US Joint all Domain Command and Control System, the Russian “Sozvezdie ACS”, China’s “Integrated Command Platform” and the Elbit Torch2H C2 System. For an illustration of such a principle in practice, see: Congressional Research Service (CRS), “Joint All-Domain Command and Control (JADC2)”, CRS, Washington, 21 January 2022.
- 74 “Palantir AIP”, Palantir: <https://www.palantir.com/platforms/aip/>.

In their most advanced and integrated forms, such assemblages of DSS capabilities would theoretically automate an entire chain of “decisions” leading to the use of force, such that the human’s role is reduced to a single decision: either approving or negating a proposed plan for the use of force.⁷⁵

75 R. S. Cohen, “Goldfein: USAF, Navy Experimenting with Multi-Domain Operations”, Air Force Magazine, 15 October 2019: <https://www.airforcemag.com/goldfein-usaf-navy-experimenting-with-multi-domain-operations/>; J.A.P. Smallegange *et al.*, *Big Data and Artificial Intelligence for Decision Making: Dutch Position Paper*, NATO Science & Technology Organization, Brussels, STO-MP-IST-160, 2018. See also the Rafael Fire Weaver System, A. Holland Michel, “Inside the messy ethics of making war with machines”, *MIT Technology Review*, September/October 2023: <https://www.technologyreview.com/2023/08/16/1077386/war-machines/>. Though as one expert noted, military personnel involved in the use of force continue to have specific roles that are only relevant, in many cases, to one decision support role; interview with Maria Riveiro, November 2021.

SECTION 4

UNCERTAINTIES, ASSUMPTIONS AND BIASES IN DECISION SUPPORT SYSTEMS

The use of DSS can and does result in failures, sometimes regularly,⁷⁶ for a wide variety of reasons. These reasons include limitations to the DSS itself and challenges in the interaction between DSS and human users. The use of increasingly complex DSS will also result in failures, in spite of the expected advances outlined in the previous sections. If human operators are not adequately aware of the possibility and the causes of such failures, their decisions are more likely to be sub-optimal and result in unintended harm⁷⁷ and there is a higher likelihood that they will trust (or mistrust) the output of a system when they should not.⁷⁸

Human operators must therefore have the capacity to identify erroneous DSS outputs or to account for the possibility that a given system may generate a sub-optimal or inappropriate solution in any given context.⁷⁹ Human users must also know how to exercise appropriate legal judgement when making a decision to use force with the support of a DSS; this is important because even “correct” outputs could support incorrect decisions (a target detection system might correctly detect a soldier from an adversary force, but deciding to kill that individual might nevertheless be illegal if they are *hors de combat*).

In other words, humans must maintain the capacity to make the right decision *regardless* of whether their DSS is correct or incorrect.

This capacity relies on the human user’s ability to gauge the uncertainties, assumptions and biases that are embedded in those DSS outputs in relation to the unique context of the decisions they support. These factors grow and become more challenging as the systems become more complex, particularly when they incorporate machine learning and when they are used for a wider variety of less mathematically definable tasks.

⁷⁶ For example, even though automatic target recognition is often described as a proven technology, it is still beset with serious challenges. See S.K. Rogers *et al.*, “The life and death of ATR/Sensor Fusion and the hope for resurrection”, in F.A. Sadjadi and A. Mahalanobis (eds.), *Automatic Target Recognition XVIII*, Proc. of SPIE Vol. 6967, 696702, 2008; B.J. Schachter, *Automatic Target Recognition*, Vol. 4, SPIE, Bellingham, 2020; A.J. Reiner, J.G. Hollands and G.A. Jamieson, “Target Detection and Identification Performance Using an Automatic Target Detection System”, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 59, Issue 2, 2017, pp. 242–258.

⁷⁷ J. Pfautz *et al.*, “The Role of Meta-Information in C2 Decision-Support Systems”, 2006 Command and Control Research and Technology Symposium, San Diego. For background on how analysts interact with uncertainty in mapping systems, see A.M. MacEachren *et al.*, “Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know”, *Cartography and Geographic Information Science*, Vol. 32, Issue 3, 2005, pp. 140–143.

⁷⁸ M. Riveiro *et al.*, “Effects of visualizing uncertainty on decision-making in a target identification scenario”, *Computers & Graphics*, Vol. 41, June 2014.

⁷⁹ This was illustrated most famously in the case of the Russian military officer Stanislav Petrov, whose decision to disregard the output of a threat detection system erroneously showing a volley of incoming missiles helped prevent a retaliation attack that likely would have precipitated a nuclear war with the West.

4.1 UNCERTAINTIES

Every DSS output, even those from a ‘traditional’ DSS that does not incorporate machine learning, is beset by intersecting forms of uncertainty as to whether it accurately represents the information that it purports to represent or whether it reflects the optimal solution to a given problem.⁸⁰ Uncertainty can be challenging to convey to the human user, whose capacity for making sound decisions is itself a source of uncertainty. Certain critical uncertainties are greater and more challenging to convey in machine learning-based DSS.



Some of the fundamental sources of uncertainty in DSS outputs in general include:

- Uncertainty as to the accuracy or representativeness of the data inputs that the DSS use.⁸¹ All data sources can contain errors or inaccuracies.⁸² Hard data such as imagery or electromagnetic signals can be degraded as a result of environmental factors, for example. Human-generated data such as intelligence reports can contain mistakes. Similarly, if a particular type of sensor can't detect hidden objects, the absence of those objects as reflected in the data cannot be

⁸⁰ H. Atoyan, J.-M. Robert and J.-R. Duquet, “Uncertainties in complex dynamic environments”, *Journal d’Interaction Personne-Système*, Vol. 2, No. 1, Art. 5, January 2011, pp. 3–5. For a granular list of types of uncertainty, see: J. Chung and S. Wark, *Visualising Uncertainty for Decision Support*, Australian Department of Defense, Defence Science and Technology Group, Fishermans Bend, 2016.

⁸¹ Interview with Maria Riveiro, November 2021.

⁸² For example, the Global Navigation Satellite System data that can be used to mark the location of targets is only accurate to a general area of varying size. In any GNSS reading there is some degree of uncertainty as to the object’s exact location relative to the indicated location; when such data are used as the basis for a strike, there would be uncertainty as to whether the intended target would actually be hit and whether unintended targets would be affected (which would be compounded by uncertainty regarding the precision of the warhead). National Coordination Office for Space-Based Positioning, Navigation, and Timing, “GPS Accuracy”, GPS.gov, 2021: <https://www.gps.gov/systems/gps/performance/accuracy/#speed>. During U.S. targeting operations in Iraq in the early stages of the Iraq War, the GPS coordinates that were often used to locate targets for strikes were only accurate to a radius of 100 meters from the target. *Off Target: The Conduct of the War and Civilian Casualties in Iraq*, Human Rights Watch, 2003.

taken as proof that they are not present. The fusion of disparate data sources, each of which have their own inherent inaccuracies, can expand the total uncertainty of the output.⁸³

- Uncertainty as to the integrity or veracity of data. The data that a DSS processes may include incorrect or corrupted datapoints. This is especially problematic in contexts where adversaries might intentionally compromise that data (say, by broadcasting fake radio signals or by generating fake online media).⁸⁴ Systems that rely on live communication connections could be vulnerable to jamming or spoofing that might similarly compromise the integrity or veracity of the system's data input.⁸⁵
- Uncertainty as to the system's scripting or calibration.⁸⁶ Many types of DSS must be calibrated to the particular context in which they are used, and/or scripted with information or parameters to generate the desired output. If this scripting or calibration includes mistakes or is inappropriate relative to the context, the system may produce a suboptimal "solution." For example, a human analyst may log the wrong number or type of targets in the input for a weapons target assignment system or script the wrong priority value to a particular objective.⁸⁷ A planning tool might be scripted to optimize speed above all other factors, but if there are civilians present, it would be more appropriate for the system to optimize reducing risk of harm to civilians. *These issues in the system's model often derive from inappropriate or incorrect assumptions about the environment, entities or problems that the model is supposed to represent (see next section).*
- Uncertainty relating to the DSS interface. Even when the system data and model are correct, the computer interfaces of DSS may fail to correctly convey critical information to the user because of how the output is displayed, which can cause the user to draw erroneous conclusions from the DSS output.⁸⁸
- Predictive uncertainty. When a system is "predicting" an unknown value (for example, predicting whether an adversary is likely to take Route A or Route B in response to an attack or "predicting" that an armed vehicle in an image is likely to fire), there is always uncertainty as to whether the prediction will be "correct." This uncertainty is often quantified as a "confidence score" or a probability reflecting the degree to which the thing being predicted

⁸³ For an overview in uncertainty in fusion, see K. Rein, *Aspects of Uncertainty in Soft Data Fusion*, NATO Science & Technology Organization, Brussels, STO-EN-IST-134; B. Khaleghi *et al.*, "Multisensor data fusion: A review of the state-of-the-art", *Information Fusion*, Vol. 14, Issue 1, 2013. See also: A. Pang, "Visualizing Uncertainty in Geo-spatial Data", paper prepared for a committee of the Computer Science and Telecommunications Board, 2001.

⁸⁴ J. Chung and S. Wark, *Visualising Uncertainty for Decision Support*, Australian Department of Defense, Defence Science and Technology Group, Fishermans Bend, 2016; A.M. MacEachren *et al.*, "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know", *Cartography and Geographic Information Science*, Vol. 32, Issue 3, 2005, p. 146. For an overview of these kinds of data issues, see: A. Holland Michel, *Known Unknowns: Data Issues and Military Autonomous Systems*, United Nations Institute for Disarmament Research, Geneva, 2021; see also: H. Atoyan, J.-M. Robert and J.-R. Duquet, "Uncertainties in complex dynamic environments", *Journal d'Interaction Personne-Système*, Vol. 2, No. 1, Art. 5, January 2011.

⁸⁵ Interview with Margarita Konaev, October 2021; interview with Milind Kulshreshtha, September 2021; S. Waterman, "Probing the Fragility of JADC2", *SIGNAL*, 1 August 2021: <https://web.archive.org/web/20210802054354/https://www.afcea.org/content/probing-fragility-jadc2>; J. Simpson, "Operations in deception: corrupting the sensing grid of the enemy", *The Forge*, 2021.

⁸⁶ Interview with Jennifer Rooke, January 2022 (via email). For examples of these kinds of scripting errors in the context of medical DSS, see: A. Wright *et al.*, "Analysis of clinical decision support system malfunctions: a case series and survey", *Journal of the American Medical Informatics Association*, 23(6), 2016.

⁸⁷ In the case of the erroneous strike on a Doctors Without Borders hospital in Kunduz, Afghanistan in 2015, the failure of human operators to load a "no strike list" of non-military facilities into an attack aircraft's guidance system contributed to the misidentification of the facility as a Taliban compound. Further information about this system and its limitations was redacted in the publicly released version of the official report on the incident. "Investigation Report of the Airstrike on the Médecins Sans Frontières / Doctors Without Borders Trauma Center in Kunduz, Afghanistan on 3 October 2015", United States Central Command, pp. 52 and 73.

⁸⁸ Confusing interface designs in air defence systems were found to have contributed to human decisions that resulted in the shootdown of Iran Air 655 in 1988 and two friendly fire shootdowns in Iraq in 2004. M.L. Cummings, "Automation and Accountability in Decision Support System Interface Design", *The Journal of Technology Studies*, 1 January 2006;

matches previously observed patterns. These scores are an imperfect metric of whether that thing will or will not be true in the given case.

- Uncertainty as to the completeness of the system's model or data.⁸⁹ If a system's data on the context it is representing or model of the problem it is solving are excessively incomplete⁹⁰ or inappropriately abstract, and if these shortfalls are not taken into consideration by the decision maker (say, by seeking out additional information to fill those gaps or secondary analyses to validate the system's output), human decisions based on those outputs could be erroneous or misaligned with legal requirements.⁹¹ *In a complex, uncontrolled environment, no DSS can ever account for every single factor in its operational space and no system model can be a perfect representation of the real world condition or process that it is modelling⁹² or simulating.⁹³* Therefore, this is a persistent source of uncertainty.
- Uncertainty as to the DSS user's fitness to make the right decision based on the system's output.⁹⁴ For example, if a user is tired, distressed or prone to particular cognitive biases, they will be less likely to make the "right" decision or an "optimal" decision on the basis of a DSS or apply legally mandated human judgement.

-
- 89 H. Atoyan, J.-M. Robert and J.-R. Duquet, "Uncertainties in complex dynamic environments", *Journal d'Interaction Personne-Système*, Vol. 2, No. 1, Art. 5, January 2011, p. 15; A.M. MacEachren *et al.*, "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know", *Cartography and Geographic Information Science*, Vol. 32, Issue 3, 2005, p. 143.
- 90 For instance, a target recognition system that identifies an approaching airplane might be too incomplete to serve as the basis for a shootdown decision if it does not take into account the transponder alerts from that aircraft. I. Bode and T. Watts, *Meaning-less Human Control: Lessons from air defence systems on meaningful human control for the debate on Autonomous Weapon System*, Drone Wars and the Centre for War Studies, University of Southern Denmark, 2021, pp. 46–51. A course-of-action planner will be inappropriate for outdoor operations if it does not account for weather phenomena that could have a significant impact on the likelihood of success of a given plan or of the precision of the weapons considered for use. An intelligence preparation of the battlefield tool will be incomplete if it lacks relevant information about, say, civilian infrastructure or if it does not capture the full spectrum of passable routes that a military entity can take to reach a target. See also: N.M. de Reus, P.J.M. Kerbusch and M.P.D. Schadd, *Geospatial analysis for Machine Learning in Tactical Decision Support*, NATO Science & Technology Organization, Brussels, 2021, MP-MSG-184-08.
- 91 M.L. Cummings, "Automation and Accountability in Decision Support System Interface Design", *The Journal of Technology Studies*, 1 January 2006, p. 24. Failing to take into account the unique structural characteristics of a target, for instance, could undermine the accuracy of the collateral damage model of that target. Attributing the same performance characteristics to every adversary target will undermine weapons target assignment calculations. A simulation tool will be inaccurate if it only has the capacity to model the behaviors of 50 enemy vehicles simultaneously when in reality there are 100 vehicles in the battlespace. F.A. Maestas and L.A. Young, "Weapon effectiveness models: are they appropriate for use in force protection analyses?", *WIT Transactions on The Built Environment*, Vol. 82, 2005. "Армию направят в виртуальный мир", *Коммерсанте*, 6 December 2021: <https://www.kommersant.ru/doc/5116316>.
- 92 A map might indicate the presence of a river but it might not be able to indicate the degree to which that river is or is not passable. It might indicate the location of an enemy checkpoint but it might not be able to indicate the level of training the soldiers inside it have received. The values that reflect the "priority" or "threat" level of a potential target are abstractions of the true degree to which that entity is a threat or priority in relation to others. S. J. Banks, "Lifting Off of the Digital Plateau With Military Decision Support Systems", Master's Thesis, School of Advanced Military Studies, United States Army Command and General Staff College, 2013, pp. 43–46; D. Schmorow *et al.*, "Applied Use of Socio-Cultural Behavior Modeling and Simulation: An Emerging Challenge for C2", *Proceedings of the 14th International Command and Control Research and Technology Symposium*, June 2009, Washington.
- 93 For example, simulations of complex phenomena and events, such as a firefight in a dense urban environment, will necessarily have to rely on abstractions of certain parameters, dynamics and events. These systems cannot model the unique, individual decision-making process of every single soldier and civilian who will be implicated in the events that will unfold. Interview with Herman le Roux, November 2021.
- 94 J. Chung and S. Wark, *Visualising Uncertainty for Decision Support*, Australian Department of Defense, Defence Science and Technology Group, Fishermans Bend, 2016; H. Atoyan, J.-M. Robert and J.-R. Duquet, "Uncertainties in complex dynamic environments", *Journal d'Interaction Personne-Système*, Vol. 2, No. 1, Art. 5, January 2011.

- Machine learning-based DSS introduce additional unique uncertainties. Even high-performing machine learning systems are prone to fail when encountering inputs that do not conform to the conditions and characteristics of the data on which they were trained and tested.⁹⁵ These inputs do not have to be widely divergent to cause failures; even subtle differences in how the data are collected can be enough to generate errors.⁹⁶ And because these training sets tend to be so large, especially in the case of systems like large language models, there may be uncertainty as to the content of the datasets, making it difficult to anticipate precisely what might cause such systems to fail or when such failure might happen.⁹⁷ Furthermore, machine-learning systems may be more likely to fail in unpredictable ways, given that systems that are more descriptive or prescriptive⁹⁸ have a greater diversity of ways in which they can fail.

UNCERTAINTIES AND HUMAN-MACHINE INTERACTION

Decision makers must account for all of these uncertainties when using a DSS. However, uncertainties can be very difficult to convey precisely or accurately. Different types of uncertainty cannot be quantified or estimated using a single metric or indicator (for example, uncertainty about an object's location is measured differently than uncertainty regarding that object's velocity). Some sources of uncertainty may be very difficult or simply impossible to convey; uncertainty as to whether an event will or will not happen in the future is particularly hard to estimate and communicate, since any prediction is only a statistical estimate.⁹⁹ This is also the case for uncertainty about whether something exists, despite not being visible in the data.

Even for individual sources of uncertainty, there are generally no standards or universally accepted guidelines as to how these could or should be characterized and conveyed to the human who must contend with them.¹⁰⁰

Uncertainty relating to the decision maker's capacity to use DSS appropriately is also not always easy to detect, characterize or quantify, especially for human operators or teams of human operators who need to detect when *they themselves* are experiencing an issue (for example, a bias that leads them to over- or under-trust the output of a system).¹⁰¹

Furthermore, when these sources of uncertainty interact, the total uncertainty grows significantly and becomes more difficult to characterize with single indicators or metrics that can be effectively conveyed to the human.¹⁰² The presentation of model uncertainty has been an ongoing research challenge for decades; meanwhile, efforts to develop "explainability" features that

95 M.L. Cummings, "The Surprising Brittleness of AI", Women Corporate Directors; A.J. Lohn, "Estimating the Brittleness of AI: Safety Integrity Levels and the Need for Testing Out-Of-Distribution Performance", 2020: <https://arxiv.org/abs/2009.00802>.

96 For example, P. Tucker, "This Air Force Targeting AI Thought It Had a 90% Success Rate. It Was More Like 25%", Defense One, 9 December 2021: <https://www.defenseone.com/technology/2021/12/air-force-targeting-ai-thought-it-had-90-success-rate-it-was-more-25/187437/>.

97 A. Holland Michel, *Known Unknowns: Data Issues and Military Autonomous Systems*, United Nations Institute for Disarmament Research, Geneva, 2021; M.M. Maas, "Regulating for 'Normal AI Accidents' — Operational Lessons for the Responsible Governance of AI Deployment", in *Proceedings of the 2018 AAAI / ACM Conference on Artificial Intelligence, Ethics and Society*, New Orleans; R.V. Yampolskiy, "Unpredictability of AI", 2019: <https://arxiv.org/abs/1905.13053>.

98 For example, systems that don't just detect targets but also characterize their "threat level" or systems that not only prescribe a detailed COA but also predict how an adversary will respond to it.

99 Interview with Maria Riveiro, November 2021; J. Quiñonero-Candela *et al.*, "Evaluating Predictive Uncertainty Challenge", in J. Quiñonero-Candela *et al.* (eds), *MLCW 2005*, LNAI 3944.

100 A.M. MacEachren *et al.*, "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know", *Cartography and Geographic Information Science*, Vol. 32, Issue 3, 2005.

101 K. Okamura and S. Yamada, "Adaptive trust calibration for human-AI collaboration", *PLoS ONE*, 15 (2), 2020: e0229132.

102 Similarly, total uncertainty will grow in scale and complexity if a single DSS is incorporating multiple decision support roles in a single output (as discussed on page 30). Interview with Maria Riveiro, November 2021; L.C. Dias, C.H. Antunes and D.R. Insua, "Dealing with uncertainty in Decision Support Systems: Recent trends 2000--2011", *Intelligent Decision Technologies*, Vol. 6, Issue 4, October 2012.

“explain” the outputs of machine learning-based systems in a simple understandable manner have struggled to balance the need to be understandable with the risk of being overly reductive.¹⁰³

4.2 ASSUMPTIONS

While the material characteristics¹⁰⁴ of objects or phenomena may be measurable, albeit with some uncertainty, the *meaning of these characteristics*¹⁰⁵ cannot be absolutely defined mathematically. Therefore, DSS are built around assumptions that assign meaning to mathematically definable attributes.¹⁰⁶ This is sometimes described as the process of turning “data” into “information.”

If these assumptions are misaligned with a reality they purport to represent, they can lead to DSS outputs that contribute to decisions resulting in unintended or unlawful harm. As systems become more “automated” they rely more heavily on a greater number of assumptions, some of which would have previously fallen to the judgement of the human user. In these cases, it becomes more difficult for the human user to account for these computerized assumptions when assessing an output.

ASSUMPTIONS THAT TURN DATA INTO INFORMATION

Data on their own are meaningless. Assumptions are necessary to turn such data into actionable information.

For example, the number of communications between two individual mobile phones is measurable and mathematically unambiguous; however, establishing whether these communications indicate an organizational relationship between two people is based on an assumption about the meaning of the number of communications and an assumption that they are the sole users of those phones.¹⁰⁷ Estimating an observed individual’s “intent” in order to decide whether they can be legitimately targeted for attack hinges on assumptions about the indicators of intent, since intent itself cannot be measured and can only be induced from measurable proxy features.¹⁰⁸

¹⁰³ R. V. Yampolskiy, “Unexplainability and Incomprehensibility of Artificial Intelligence”, 2019: <https://arxiv.org/abs/1907.03869>; R. Schmelzer, “Understanding Explainable AI”, Forbes, 23 July 2019: <https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/?sh=1292ca7b7c9e>.

¹⁰⁴ Size, number, shape, time, location, velocity, etc.

¹⁰⁵ For example, type of object (e.g. “car” or “truck”, “adult” or “child”), the identity of the object or individual (“military truck” or “medical truck”, “soldier” or “journalist”), purpose of the object or individual (“truck on a resupply mission,” “soldier retreating”), the status of the object or individual (“neutralized truck,” “incapacitated soldier”), or the relationship of that object or individual to other entities (“truck in a convoy with another vehicle,” “soldier’s commanding officer”).

¹⁰⁶ S.J. Banks, “Lifting Off of the Digital Plateau With Military Decision Support Systems”, Master’s Thesis, School of Advanced Military Studies, United States Army Command and General Staff College, 2013, pp. 43–46. For a definition of “assumptions” and how they differ from facts in the context of military planning, see, for example: FM 6–0 Commander and Staff Organization and Operations, “FM 6–0: Commander and Staff Organization and Operations,” Headquarters, Department of the U.S. Army, May 2014, p. 4–2.

¹⁰⁷ M. Robbins, “Has a rampaging AI algorithm really killed thousands in Pakistan?”, The Guardian, 18 February 2016: <https://www.theguardian.com/science/the-lay-scientist/2016/feb/18/has-a-rampaging-ai-algorithm-really-killed-thousands-in-pakistan>; P.V. Fellman and R. Wright, “Modeling Terrorist Networks, Complex Systems at the Mid-range”, 2014: <https://arxiv.org/abs/1405.6989>.

¹⁰⁸ H. Irandoust and A. Benaskeur, “Human-Autonomy Teaming for Critical Command and Control Functions”, Defence Research and Development Canada, Ottawa, 2020, p. 5; J.N. Roux and J.H. van Vuuren, “Threat evaluation and weapon assignment decision support: A review of the state of the art”, *ORION*, Vol. 23(2), 2007; interview with Maria Riveiro, November 2021; (That is, short of asking the individual, “What is your intent?”).



An analytical system may be capable of identifying a strong correlation between events by mathematical means, but establishing causation (i.e., establishing the meaning of those correlations) is inherently a non-mathematical process.¹⁰⁹

Models that generate solutions to “planning problems” likewise rely on assumptions. For example, a route-planning algorithm might operate on the assumption that a shorter route is preferable to a longer route. Optimization relies on the assignment of numerical values to abstract qualities like the military significance of a target or the vulnerability of an asset¹¹⁰ so that these can be prioritized or deprioritized accordingly.

Assumptions are also necessary in simulations. A simulation might rely on game-theoretic assumptions – say, that an enemy will prioritize tactics that reduce risk of harm to its own forces – as well as assumptions about whether given variables (such as weather) are likely to change or remain the same.

All DSS rely on assumptions. But some rely on more assumptions than others. DSS that extract more information from data and engage in more complex planning or prediction generally embed a greater number of assumptions than simpler, single-task systems.

¹⁰⁹ H. Irandoust and A. Benaskeur, “Human–Autonomy Teaming for Critical Command and Control Functions”, Defence Research and Development Canada, Ottawa, 2020; for example, machine analysis system that identifies an object as a “threat” is, in fact, merely noting that that object exhibits characteristics associated with previous threats. O. Daniels, “Speeding Up the OODA Loop with AI A Helpful or Limiting Framework?”, Joint Air & Space Power Conference 2021; A. Deeks, N. Lubell and D. Murray, “Machine Learning, Artificial Intelligence, and the Use of Force by States”, *Journal of National Security Law & Policy*, Vol 10:1, p. 12.

¹¹⁰ D. Pedersen et al., “Decision Support System Engineering for Time Critical Targeting”, MITRE Technical Paper, Bedford, 1999, p. 5.

CLEAR ASSUMPTIONS VS. UNCLEAR ASSUMPTIONS

These assumptions are sometimes clear and known. For example, when a human “scripts” (codes) a system on the basis of various clear and deliberate assumptions, these assumptions are evident and unambiguous. But other times, assumptions are neither clear nor known.

For example, an “assumption” might take the form of an abstract statistical principle buried deep within a system’s formula for “predicting” an unknown value. Many models rely on multiple layers of interaction or aggregated assumptions. An enemy vehicle may be more likely to be marked as hostile if it is exhibiting driving behaviours that diverge from a previously established pattern of supposedly “normal” driving behaviour *and if* it is travelling from a particular area *and if* it is observed shortly after a nearby military action by the adversary groups.¹¹¹ A simulation that anticipates how an enemy force is likely to respond to a given action will be based on both assumptions about its observed actions up to that moment, as well as assumptions about how it might build upon those actions in its next moves.

As the number of assumptions embedded in a system goes up (and the number of steps in which human judgement is considered necessary for making a decision on the basis of a DSS output goes down), it becomes harder for humans to know and account for all of these assumptions in their decision.

HUMAN VS. COMPUTATIONAL ASSUMPTIONS

The process of gleaning meaning from information by way of assumptions is, of course, not unique to computerized DSS. It is a fundamental element of all military intelligence analysis and decision-making. However, there is a fundamental difference between the assumptions that underpin human reasoning and the assumptions coded into a DSS’s model.

Human analytical processes are not mathematical or computational. They involve the application of logic, the interpretation and application of moral imperatives, contextual understanding and judgement. Therefore, DSS can only offer a proxy of human reasoning. For example, systems that are supposedly “hardwired” with legal requirements such as distinction and proportionality cannot encode the process by which a human decision maker makes legal judgements. Instead, they are based on a mathematical approximation of selected measurable phenomena and hard-wired assumptions about their relationships.¹¹² The notion that relying on the “advice” of such systems can satisfy legal requirements remains a heavily contested proposition.

In addition to assumptions embedded in the DSS, when a human uses a DSS output as the basis for a decision, they are very likely to rely on their own assumptions about the DSS and the context. For example, if a DSS only indicates the location of features on a map, it falls to the human to leverage assumptions about the significance of those features in relation to their objective. Similarly, a fairly simple analytical DSS tool might only detect that a vehicle is travelling faster than any vehicles in its vicinity, leaving it to the human user to make an assumption as to whether the fact that the vehicle is speeding is likely to make it a threat.

¹¹¹ For an example of how such assumptions serve as the basis for targeting decisions in non-computerized analysis, see: “Testimony of a French drone operator: anticipatory strikes in the Sahel”, European Forum on Armed Drones, 16 February 2022: <https://www.efadrones.org/testimony-of-a-french-drone-operator-anticipatory-strikes-in-the-sahel/>.

¹¹² K. Klonowska, “Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare”, Asser Research Paper 2021-02, p. 18.

With each additional task that a DSS carries out, it takes on more assumptions that would have previously fallen to the human decision maker. The more “automated” DSS functions become, the harder it will be to make these assumptions available to the user and to validate them at either the time of development, in testing or during use.

More fundamentally, by transferring an assumption that is a fundamental aspect of a human decision from the human to a machine, human responsibility for that decision is potentially diminished.

PROBLEMATIC ASSUMPTIONS

Just like all analytical assumptions, the assumptions embedded in a DSS can be problematic. Consider, for example, a model that supports the assessment of whether an individual is taking a direct part in hostilities.¹¹³ The system may predict with high confidence that an individual is “carrying a weapon” but not all people necessarily carry weapons to cause harm to one party to the conflict in support of another. And even in an area of active hostilities, the detection of a weapon on its own does not make the person carrying it a lawful target for attack (to say nothing of the fact that not all objects that resemble weapons, either to a computer vision system or a human, are actually weapons¹¹⁴). For such a system’s output to serve as the basis for a targeting decision, it therefore has to be complemented by other supporting evidence and countervailing assumptions.

Similarly, an individual may exhibit the same measurable behaviours as known enemy combatants (such as visiting the same locations or speaking with the same people) but engage in those behaviours for entirely different reasons. For example, a person may engage in those activities because they work as a war correspondent.¹¹⁵ Any system that labels such an individual as being a likely combatant therefore relies on an assumption that misrepresents the reality on the ground.

¹¹³ N. Melzer, “Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law”, ICRC, Geneva, 2009.

¹¹⁴ Consider, for example, the case where decision makers approved an airstrike on civilians in Gaza in December 2008 after mistaking the oxygen tanks that they were carrying for rockets. “Precisely Wrong: Gaza Civilians Killed by Israeli Drone-Launched Missiles”, Human Rights Watch, 30 June 2009: <https://www.hrw.org/report/2009/06/30/precisely-wrong/gaza-civilians-killed-israeli-drone-launched-missiles>.

¹¹⁵ One U.S. data analytics system, SKYNET, identified a well-known journalist as having a high probability of being a member of Al Qaeda, based on his location and cellphone usage. Indeed, the system identified him as having a higher probability than many *actual* members of the terror group. M. Robbins, “Has a rampaging AI algorithm really killed thousands in Pakistan?”, The Guardian, 18 February 2016: <https://www.theguardian.com/science/the-lay-scientist/2016/feb/18/has-a-rampaging-ai-algorithm-really-killed-thousands-in-pakistan>; K. Klonowska, “Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare”, Asser Research Paper 2021-02.

Likewise, all systems that “predict” future events rely on a fundamental assumption that the same factors that lead to (or that were merely correlated with) those events in the past will do so again in the future – an inevitably limited means of actually determining what will happen next.

In planning systems, assumptions can also be problematic. A system might run on an assumption that “what is bad for you is good for your opponent, and vice-versa,” when in reality, some outcomes might be bad for both parties to a conflict.¹¹⁶ In some types of military applications, the metrics of “success”¹¹⁷ that a DSS might use in order to recommend an action with the greatest “probability of success” are not necessarily anchored to what success and complying with the law would actually mean in the real world.¹¹⁸

ASSUMPTIONS IN MACHINE LEARNING-BASED DECISION SUPPORT SYSTEMS

The complexity and knowability of these various assumptions grow further with the advent of non-deterministic systems such as machine learning-based DSS. For example, a deterministic missile warning tool will “identify” missiles based on a relatively small number of criteria such as size, altitude, speed and trajectory, and, by extension, will generate warnings on the basis of a static, clearly defined and enumerated set of human-scripted assumptions about the “meaning” of those detected characteristics. By contrast, a machine learning-based system might identify missiles on the basis of the statistical degree to which an object matches the thousands or millions of “missiles” in that system’s training data. In a sense, such systems “script” themselves, and each of the millions of parameters in such a model represents an “assumption” about the meaning of the observed object or phenomena’s characteristics. Such systems may automatically embed certain assumptions that could result in unintended or unlawful harm and diverge from the wishes of those developing or deploying them.¹¹⁹

4.3 BIASES

Decision support tools can be prone to a range of technical biases. Biases arise when there are significant and systematic differences between the context that the system encounters in use and the environment for which it was designed and tested (or, in the case of machine learning systems, the data on which the system was trained.)

Biases can cause systems to have a systematic propensity to exhibit worse or more harmful performance in response to some types of inputs or subjects than others. If human users of DSS are not aware of these biases and do not take steps to account for them, their decisions are liable to be disproportionately more harmful in relation to the objects or entities towards which the system is biased.

¹¹⁶ S.N. Hamilton and W.L. Hamilton, “Adversary Modeling and Simulation in Cyber Warfare”, *Proceedings of The 23rd International Information Security Conference*, 2008.

¹¹⁷ For example, “serious/lethal wounds to standing personnel from primary warhead fragmentation or debris”.

¹¹⁸ For example, securing an area while minimizing harm to civilians and damage to civilian objects, and complying with all other applicable national and international laws. K. McKendrick, *The Application of Artificial Intelligence in Operations Planning*, NATO Science & Technology Organization, Brussels, 2017, STO-MP-SAS-OCS-ORA-2017, p. 7; S.C. Gordon, “Decision Support Tools for Warfighters”, 2000 Command and Control Research and Technology Symposium, Monterrey, p. 8. In a conventional battle, it might be possible to codify success as destroying the highest proportion of enemy vehicles while incurring the lowest proportion of losses among one’s own forces (though even this is a simplification). But in a complex counterinsurgency campaign, for example, criteria for success cannot be mathematically quantified. See also P.K. Davis, J. Kulick and M. Egner, *Implications of Modern Decision Science for Military Decision-Support Systems*, RAND Corporation, Santa Monica, 2005, p. 46.

¹¹⁹ Consider, for example, the case of an Amazon decision support system that vetted the resumes of job applicants; because the system was trained on historical data in which men were more likely to be hired for jobs than women, the system downgraded resumes that included words related to women. J. Dastin, “Amazon scraps secret AI recruiting tool that showed bias against women”, Reuters, 11 October 2018: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.

SOURCES OF BIAS

In some cases, these biases arise simply because the system encounters something that it was not designed for. If an automatic target recognition system is designed and calibrated and vetted for use against a specific adversary operating according to a specific set of tactics with a specific type of weapon, that system will likely exhibit systematically worse performance in a scenario where the adversary uses different tactics or different weapons,¹²⁰ or against a completely different military force.

In other cases, systems reflect biases that were deliberately or unknowingly coded into their architecture by those who built them. For example, if a system is coded to identify any person carrying a weapon as a potential target, this system could exhibit harmful bias against individuals who are more likely to carry weapons for non-military reasons (such as hunters or police officers) or individuals (such as photographers) who are more likely to carry objects that could be easily misidentified by sensors as weapons.

Bias is thought to be especially preponderant and more challenging to detect and mitigate in machine learning-based systems. This is due to the inherently statistical nature of their models, the propensity of historical data to reflect societal biases, the challenges of aligning system training data with the statistical characteristics of the environment to which the system is deployed, and the large volume and complexity of these datasets.

Some of the most problematic technical biases can be traced to cognitive or societal biases in the people and organizations that develop the system. Similar biases may stem from the data upon which the systems were developed and tested. These data may be skewed as a result of historical inequalities. These biases can lead to disproportionate harm to certain demographic groups.¹²¹ For example, if the team developing a tool to simulate the behaviour of a population possesses poor or misinformed cultural knowledge of that group, it is likely that the resulting simulation will be prone to misrepresenting the population.

Similarly, a key assumption in a model may embed a social bias regarding a certain group. For example, a model might operate on the assumption that individuals with certain physical characteristics are more “suspicious,” reflecting the system designer’s biased views on that group. In predictive systems, the “proxy” features (age, ethnicity, gender, etc.) used for categorizing people are derived from historical datasets that may reflect disproportionately on certain demographic groups, leading to a disproportionately high misclassification of and focus on those groups. If new targeting data resulting from the use of these biases systems are used, in turn, for further predictions, the effect of this bias will be amplified.¹²²

Furthermore, the technical biases inherent to a system¹²³ can interact with the biases of the users. This can create a feedback loop that exacerbates all of these biases’ individual effects. For example, a search or synthesis tool that has a propensity to display more threat information about

¹²⁰ S.J. Freedberg Jr., “Artificial Stupidity: Fumbling The Handoff From AI To Human Control”, *Breaking Defense*, 5 June 2017: <https://breakingdefense.com/2017/06/artificial-stupidity-fumbling-the-handoff/>; A. Deeks, “Predicting Enemies”, *Virginia Public Law and Legal Theory Research Paper No. 2018–21*, 1 March 2018, p. 1564. For this reason, one expert recommended that any time a DSS is deployed to a new environment, users must assess whether it will perform with the same reliability as previous employments or if it must be tweaked or used in a different way.

¹²¹ K. Lum and W. Isaac, “To Predict and Serve?”, *Significance*, Vol. 13(5); J. Buolamwini and T. Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, *Proceedings of Machine Learning Research*, 81, 2018; K. Fisher, “Artificial Intelligence, Warfare, and Bias”, *PRIO Blogs*, 6 September 2021.

¹²² A.E.R. Prince and D. Schwarcz, “Proxy Discrimination in the Age of Artificial Intelligence and Big Data”, *Iowa Law Review*, 105, 1257, 2020.

¹²³ K. Lai et al., “Assessing Risks of Biases in Cognitive Decision Support Systems”, 28th European Signal Processing Conference (EUSIPCO), Amsterdam, 2020.

a particular type of area as compared to other areas¹²⁴ would exacerbate the bias of an operator who tends to assume that a higher frequency of *displayed* threat data for an area indicates an *actual* higher threat in that area. It would also amplify the bias of an operator who has negative attitudes towards the population of those areas.

THE CHALLENGES OF MITIGATING BIAS

The military domain poses many obstacles to the debiasing of system inputs and architectures. Militaries developing DSS systems will rarely have a complete picture of the context in which those systems are going to be used. Such information is often either classified, difficult to collect or validate, subject to adversarial conditions or nonexistent. Biases can be especially hard to identify in machine learning-based DSS systems, given the inevitability of a wide range of bias types in large datasets and the fact that such biases may only become evident once the system is deployed.¹²⁵

Even when DSS can be vetted for bias by technical means, conflicts are constantly changing. Therefore, a system that did not exhibit harmful biases when it was first used could acquire biases during use. And since this “shift” is gradual, operators may not be able to detect it until after it leads to misaligned outputs or outcomes resulting in unintended or unlawful harm.¹²⁶ The issue of emergent bias could be compounded in “active learning” DSS that continually evolve during use. Even though such active learning is intended to help systems remain well-suited for their evolving environment, automatic updates to systems could also introduce unintended biases.¹²⁷

It could also be more challenging for users to counteract biases in a DSS if these biases align with their own personal biases. For example, if a user of a system has cultural biases against people who have the “proxy features” that a threat-warning system is statistically biased against (which it may hold because of societal biases), they could be less likely to counteract the system’s propensity to mark those people as “threats.” DSS biases can also interact harmfully with other human cognitive biases (see Section 4.5).

¹²⁴ For example, because more historical threat data has been generated for that area or because past users have disproportionately sought out threat data for that area. For a comprehensive study of bias in search engines, see: S. Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, New York, 2018. This phenomenon has also been widely observed in location-based predictive policing tools, whose use has resulted in disproportionate policing activity and, by extension, harms, on low-income communities. See: A. Sankin *et al.*, “Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them”, The Markup, 2 December 2021: <https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them>.

¹²⁵ S. Verma, M. Ernst and R. Just, “Removing biased data to improve fairness and accuracy”, 2021: <https://arxiv.org/abs/2102.03054>; S. Leavy, B. O’Sullivan and E. Siaper, “Data, Power and Bias in Artificial Intelligence”, 2020: <https://arxiv.org/abs/2008.07341>.

¹²⁶ S. Shendré, “Model Drift in Machine Learning”, Towards data science, 13 May 2020: <https://towardsdatascience.com/model-drift-in-machine-learning-models-8f7e7413b563>; D. Sculley *et al.*, “Machine Learning: The High Interest Credit Card of Technical Debt”, *Software Engineering for Machine Learning*, NIPS 2014 Workshop.

¹²⁷ *Ibid.*

4.4 THE VARYING RELEVANCE OF UNCERTAINTIES, ASSUMPTIONS AND BIASES

All DSS outputs are embedded with uncertainties, assumptions and biases. However, the extent to which these need to be taken into account in any decision is *highly context dependent*. A potentially problematic uncertainty, assumption or bias may only be relevant to a decision when it can result in unintended or unlawful harm or error.

Identifying uncertainties, assumptions and biases is in itself a challenge; determining whether they are likely to have a bearing on the outcome of a supported decision in any given instance can be exceedingly difficult. The factors that determine the relevance of a particular uncertainty, assumption or bias may not be readily observable or knowable¹²⁸ and the policies and protocols that govern the use of a DSS may not be able to account for every system characteristic in every potential context of use where its relevance may or may not be a matter of consideration.

Examples of Relevance and Irrelevance:

- In an environment where civilians are unlikely to be present, uncertainty about the precise location of an observed target may be far less significant than the same uncertainty in an urban setting, where a difference of just a few meters in the use of weapons could have dramatic consequences for civilians.¹²⁹
- Uncertainty about whether it may rain during an operation is only relevant in a situation where rain, for example, creates a risk of dangerous weapon malfunction – as opposed to cases where rain would have no discernible effect.¹³⁰
- The validity of the assumption that a vehicle's direction of travel is an indicator of its identity is dependent on where friendly and enemy positions are located relative to its location – information that might not always be available.
- Assuming that regular communication between two combatants indicates an organizational link may not be valid if those two individuals are also family members, friends or personal contacts unrelated to the conflict.
- Estimating the overall effects of system bias on a DSS's performance in a particular context relies on knowledge of how likely it is to encounter inputs for which it has bias.¹³¹ A system's bias to misclassify a particular type of vehicle may be irrelevant in contexts where such vehicles are not used.

Because of the sensitivity of system characteristics to context, the processes by which organizations audit complex DSS may be ill-equipped to anticipate the effects of these factors once the system is deployed. This may be especially true in the case of complex machine learning-based systems that will inevitably display previously unknown failure modes once deployed.¹³²

Furthermore, the changing dynamics of warfare might create conditions that cause previously benign system characteristics to become problematic. An adversary might, for example, develop a new tactic that renders an assumption invalid; a new cyberweapon may emerge that introduces

¹²⁸ B. Chandrasekaran, "From Optimal to Robust COAs: Challenges in Providing Integrated Decision Support for Simulation-Based COA Planning", White Paper, February 2005.

¹²⁹ S. Muhammedally, "Preparedness in urban operations: a commander's planning checklist to protect civilians", Humanitarian Law & Policy, 11 May 2021: <https://blogs.icrc.org/law-and-policy/2021/05/11/preparedness-in-urban-operations/>; "Field Manual No. 2-91.4: Intelligence Support to Urban Operations", Headquarters Department of the U.S. Army, Washington, 20 March 2008.

¹³⁰ B. Chandrasekaran, "From Optimal to Robust COAs: Challenges in Providing Integrated Decision Support for Simulation-Based COA Planning", White Paper, February 2005.

¹³¹ A.J. Reiner, J.G. Hollands and G.A. Jamieson, "Target Detection and Identification Performance Using an Automatic Target Detection System", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 59, Issue 2, 2017, pp. 242–258.

¹³² M.A. Flournoy, A. Haines and G. Chefetz, *Building Trust through Testing: Adapting DOD's Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems*, WestExec Advisors, Washington, 2020.

new uncertainty in data sources; or the statistical features of a population might evolve, giving rise to new system biases.

Even if these factors are known and even if the relevant context is observable, it might be challenging for organizations to develop policies and protocols that account for the full spectrum of potential forms of human-system error arising from such factors.¹³³ Or they may develop practices that knowingly disregard the issue. A targeting policy, for example, may provide strict guidance on how decision makers must seek senior approval if a DSS indicates the presence of a person who may be a civilian in the vicinity of a proposed target. But it might not provide guidance on how those decision makers must respond if other intelligence sources or DSS indicate with higher confidence that there are no such persons present.

The failure to recognize or account for relevant factors in the use of DSS appears to have contributed to instances of harm in recent conflicts. For example, an airstrike on a bomb factory that killed 70 civilians in Al Hawijah, Iraq in June 2015 was carried out on the basis of a CDE model that did not account for the possibility of secondary explosions from the strike.¹³⁴ The uncertainty in the output would not be so relevant if the proposed strike was targeting an individual in a sparse area. However, in this case, the strike was targeting a known vehicle-borne explosive device (VBIED) factory in a densely populated area. Therefore, in this case, the output's uncertainty was highly relevant. A military official noted during the investigation that although the CDE tool showed low collateral damage for the proposed strike, "knowledge that the target was a VBIED factory" should have made it unreasonable to assume that that CDE was accurate.¹³⁵ A secondary explosion following a strike in Mosul, Syria on 17 March 2017 was also not accounted for in the pre-strike CDE estimate; that strike resulted in 105 civilian casualties. In this case, analysts were unaware of the possible presence of secondary explosives in the targeted building, but they were aware of the presence of civilians.¹³⁶

¹³³ For a discussion of this phenomenon in the use of complex DSS in the legal domain, see D. Klutetz and D.K. Mulligan, "Automated Decision Support Technologies and the Legal Profession", 15 July 2019, *Berkeley Technology Law Journal*, forthcoming: <https://ssrn.com/abstract=3443063> or <http://dx.doi.org/10.2139/ssrn.3443063>.

¹³⁴ L. Treffers, "Newly released documents reveal Dutch knew about possible high risk to civilians at Hawijah", *Airwars*, 23 March 2020: <https://airwars.org/news-and-investigations/newly-released-documents-reveal-the-dutch-knew-about-possible-high-risk-to-civilians-at-hawijah/>; "Al Hawijah ISIL VBIED Factory strike", United States Central Command, 2015, p. 23/056, available for download at: <https://www.nytimes.com/interactive/2021/us/civilian-casualty-files.html>.

¹³⁵ "Al Hawijah ISIL VBIED Factory strike," United States Central Command, 2015, p. 29/064, available for download at: <https://www.nytimes.com/interactive/2021/us/civilian-casualty-files.html>.

¹³⁶ "Army Regulation 15-6 Investigation of the Alleged Mass Casualty Incident in the al Jadidah District", United States Central Command, May 2017, available for download at: <https://www.nytimes.com/interactive/2021/us/civilian-casualty-files.html>. See also the investigation of a secondary-explosion civilian casualty incident in Raqqa, Syria on July 11, 2015 which stated that "secondary explosions are not included in CDE methodology": "Raqqa, Syria July 11 2015 Strike Investigation", United States Central Command, p. 3, available for download at: <https://www.nytimes.com/interactive/2021/us/civilian-casualty-files.html>.

4.5 HUMAN LIMITATIONS

It can be challenging for human decision makers to properly identify, recognize and comprehend a DSS output's uncertainties, assumptions and biases, let alone grasp the significance of them in context.¹³⁷ In particular, humans are prone to cognitive limitations and biases that can hamper their capacity to make appropriate judgements on the basis of DSS, especially under time pressure.

While many of these issues have been studied extensively (though not definitively resolved) in the context of traditional decision support tools,¹³⁸ relatively few studies have explored these issues in relation to machine learning-based DSS and the novel challenges that such systems introduce with respect to predictability and understandability.¹³⁹ Even fewer studies have considered these specific, potentially novel challenges in the context of a military decision-making environment, let alone in relation to decisions on the use of force, where singular risks and legal obligations apply.

COGNITIVE CAPACITY

Decision makers engaged in armed conflict often have to balance myriad demands on their attention.¹⁴⁰ This can limit their capacity to account for complex indications of every relevant uncertainty, bias or assumption in every single instance of DSS use.¹⁴¹ In some cases where such information is necessarily rich and dense, representing these factors could actually even degrade a decision maker's overall decision-making capacity, due to information overload.¹⁴² And yet omitting such information also carries the risk of leading to an inappropriate decision. Military computer interfaces also continue to be quite rudimentary,¹⁴³ which creates further challenges for conveying these factors in a way that is both intuitive and easy to grasp while also not being overly reductive.

It can be particularly difficult for humans to make the right decision on the basis of a system's output if the time available to make that decision is limited or if the humans must juggle numerous competing demands upon their attention.¹⁴⁴ In the case of the shootdown of Iran Air 655 in 1988, stress and cognitive overload are thought to have contributed to the failure of decision makers operating the ship-borne air defence system to interpret the aircraft's radar signature, as detected by a DSS, as that of a commercial airliner.¹⁴⁵

¹³⁷ H. Langdalen, E.B. Abrahamsen and H.B. Abrahamsen, "A New Framework To Identify And Assess Hidden Assumptions In The Background Knowledge Of A Risk Assessment", *Reliability Engineering & System Safety*, Vol. 200, August 2020.

¹³⁸ Such issues, which are sometimes treated under the theme of "trust calibration" and "human-computer interaction," have been studied extensively.

¹³⁹ M. Konaev and H. Chahal, "Building trust in human-machine teams", Brookings Tech Stream, 18 February 2021: <https://www.brookings.edu/techstream/building-trust-in-human-machine-teams/>.

¹⁴⁰ J. Zhou *et al.*, "Effects of Uncertainty and Cognitive Load on User Trust in Predictive Decision Making", in R. Bernhaupt *et al.* (eds), *Human-Computer Interaction – INTERACT 2017*, Springer Cham: <https://doi.org/10.1007/978-3-319-68059-0>.

¹⁴¹ J. Chung and S. Wark, *Visualising Uncertainty for Decision Support*, Australian Department of Defense, Defence Science and Technology Group, Fishermans Bend, 2016; interview with Peter Svenmarck, November 2021; interview with Svetlana Yanushkevich, November 2021; interview with Maria Riveiro, November 2021; K. Lai *et al.*, "Assessing Risks of Biases in Cognitive Decision Support Systems", 28th European Signal Processing Conference (EUSIPCO), Amsterdam, 2020.

¹⁴² A. Pang, "Visualizing Uncertainty in Geo-spatial Data", paper prepared for a committee of the Computer Science and Telecommunications Board, 2001.

¹⁴³ Interview with Peter Svenmarck, November 2021; interview with Herman le Roux, November 2021.

¹⁴⁴ Interview with Herman le Roux, November 2021; J. Zhou *et al.*, "Effects of Uncertainty and Cognitive Load on User Trust in Predictive Decision Making", in R. Bernhaupt *et al.* (eds), *Human-Computer Interaction – INTERACT 2017*, Springer Cham: <https://doi.org/10.1007/978-3-319-68059-0>.

¹⁴⁵ This was despite the fact that operators of the same system on another ship deemed that it was not behaving in a hostile manner and despite both teams having access to information about the system's transponder emissions that identified it as a civilian craft. D. Evans, "Vincennes: A Case Study", *Naval Institute Proceedings*, Vol. 119/8/1,086, August 1993.

COGNITIVE BIASES

Humans are prone to a range of judgemental and cognitive biases that can significantly hamper their ability to contextually evaluate relevant information – including information about a decision support system’s limitations – when making a decision. Proper training can reduce the effects of certain human cognitive limitations and biases. That being said, training has been shown to have a limited effect on mitigating certain deeply rooted forms of human bias.¹⁴⁶

For example, if a DSS generates a long unbroken run of correct outputs, operators are likelier to dismiss the possibility of a subsequent output being incorrect or underestimate the probability that an uncertainty or assumption will be problematic, even when the system’s error rate is known.¹⁴⁷ Conversely, if a system generates a consistent run of incorrect outputs or even a single faulty output, the operators are more likely to disregard future outputs, even if there is still a high probability that the system will be correct in any given subsequent instance.¹⁴⁸

If a system’s outputs confirm the user’s expectations or beliefs, they are less likely to take into account relevant uncertainties, even if those uncertainties are plainly accessible.¹⁴⁹ Consider, for example, the case of the shootdown of civilian airliners in Ukraine in 2014 and Iran in 2020. In both instances, operators were likely aware that the automatic target recognition systems in their air defense weapons did not integrate data from air traffic control systems. This was an obvious and relevant source of uncertainty. And yet the knowledge of this uncertainty apparently did not dislodge their certainty that they were targeting a hostile military aircraft.¹⁵⁰

DSS can also confound users’ capacity to consider fundamental statistical principles when making a decision. For instance, if systems make certain pieces of information more available than others, humans may be more likely to estimate the frequency or probability of things related to that information to be higher, even when it’s not.¹⁵¹ Relatedly, humans have a strong tendency to assume causation when in reality the statistical evidence (and thus the statistics-based DSS output) can only indicate correlation.¹⁵²

The quality of decisions on the basis of more complex and more automated systems could also be more sensitive to the individual decision makers’ human judgemental biases, compared to decisions that require individual system task outputs to be vetted or that require additional layers of analysis and problem solving.¹⁵³

¹⁴⁶ J.E. (Hans) Korteling, J.Y.J. Gerritsma and A. Toet, “Retention and Transfer of Cognitive Bias Mitigation Interventions: A Systematic Literature Study”, *Frontiers in Psychology*, 12:629534, August 2021.

¹⁴⁷ J.D. Kulick and P.K. Davis, “Judgmental Biases in Decision Support for Strike Operations”, in A.F. Sisti and D.A. Trevisani (eds), *Enabling Technologies for Simulation Science VII, Proceedings of SPIE*, Vol. 5091, 2003, pp. 263–264.

¹⁴⁸ S. Knocton et al., “The Effect of Informing Participants of the Response Bias of an Automated Target Recognition System on Trust and Reliance Behavior”, *Human Factors*, June 2021, pp. 7–8: <https://doi.org/10.1177/00187208211021711>; L. Wang, G.A. Jamieson and J.G. Hollands, “Trust and reliance on an automated combat identification system”, *Human Factors*, 51(3), June 2009. This “under-trust” can be just as dangerous as “over-trust”: consider, for example, that a human chooses to ignore a DSS’s correct warning about the presence of persons who may be civilians in the area of a proposed strike because its previous warnings about the presence of people were incorrect. B.J. Dietvorst, J.P. Simmons and C. Massey, “Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err”, *Journal of Experimental Psychology: General*, 144 (1); V. Boulanin et al., *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*, Stockholm International Peace Research Institute, Stockholm, 2020, p. 19.

¹⁴⁹ M.L. Cummings, “Automation Bias in Intelligent Time Critical Decision Support Systems”, AIAA 1st Intelligent Systems Technical Conference, Chicago, September 2004.

¹⁵⁰ I. Bode and T. Watts, *Meaning-less Human Control: Lessons from air defence systems on meaningful human control for the debate on Autonomous Weapon System*, Drone Wars and the Centre for War Studies, University of Southern Denmark, 2021, pp. 46–49.

¹⁵¹ E. Dimara, P. Dragicevic and A. Bezerianos, “Accounting for Availability Biases in Information Visualization”, 2016: <https://arxiv.org/abs/1610.02857>.

¹⁵² B.S. Williams, “Heuristics Biases in Military Decision making”, *Military Review*, Sept–Oct 2010, p. 63.

¹⁵³ For this reason, multiple SMEs consulted for this study stressed the need for enhanced training for users of AI-enabled DSS.

It has also been widely observed that DSS can lead to dependency and complacency, which can undermine users' capacity to account for system issues or to respond appropriately when a system fails.¹⁵⁴ This effect can be especially problematic if the system is the only means by which the human can be made aware of the information relevant to their decisions.¹⁵⁵ In such cases, the use of such systems may in fact diminish operators' capacity to maintain sufficiently comprehensive contextual awareness.¹⁵⁶ The use of systems that supplant human processes of searching for information or evaluating options also reduces their grasp of the full range of potential options available to them.¹⁵⁷

In fact, in some instances it is possible that overall decision quality and consistency could be higher when operators use a system that requires some degree of manual searching, coding or validation.¹⁵⁸ That being said, operators may in some cases be more likely to accept a system's outputs uncritically if they have personally set up the system.¹⁵⁹ This is similarly problematic, since human judgemental biases and problematic assumptions in the manual setup process can cause systems to generate biased outputs.¹⁶⁰

It has also been speculated that the extensive use of DSS in decisions on the use of force may cause human agents to develop a "moral buffer" against the ethical implications of their actions.¹⁶¹ This could prevent operators from appraising their DSS outputs as rigorously as they would have if their decision were not mediated by a computerized system.¹⁶²

-
- ¹⁵⁴ Interview with Milind Kulshreshtha, September 2021; anonymous interview with a military official, October 2021; K. Goddard, A. Roudsari and J.C. Wyatt, "Automation bias: a systematic review of frequency, effect mediators, and mitigators", *Journal of the American Medical Informatics Association*, 19(1), 2012. For a canonical study on the "complacency" that can arise from the automation of certain tasks in these types of contexts, see R. Parasuraman, R. Molloy and I. L. Singh, "Performance Consequences of Automation-Induced 'Complacency'", *The International Journal of Aviation Psychology*, Vol. 3, 1993.
- ¹⁵⁵ M.L. Cummings, "Automation Bias in Intelligent Time Critical Decision Support Systems", AIAA 1st Intelligent Systems Technical Conference, Chicago, September 2004. Balancing the benefits of applying DSS to difficult tasks against the risks of "skill degradation" is regarded as a core challenge for the proper integration of DSS into new roles and critical functions. H. Atoyan, J.-M. Robert, J.-R. Duquet, "Uncertainties in complex dynamic environments", *Journal d'Interaction Personne-Système*, Vol. 2, No. 1, Art. 5, January 2011.
- ¹⁵⁶ S.J. Banks, "Lifting Off of the Digital Plateau With Military Decision Support Systems", Master's Thesis, School of Advanced Military Studies, United States Army Command and General Staff College, 2013, pp. 35–36; R. Parasuraman, T.B. Sheridan and C.D. Wickens, "A model for types and levels of human interaction with automation", *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, Vol. 30, Issue 3, May 2000, p. 291.
- ¹⁵⁷ T. Cerri *et al.*, "Using AI to Assist Commanders with Complex Decision-Making", Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC) 2018, pp. 10–11.
- ¹⁵⁸ R. Parasuraman, T.B. Sheridan and C.D. Wickens, "A model for types and levels of human interaction with automation", *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, Vol. 30, Issue 3, May 2000, p. 291; K. Goddard, A. Roudsari and J.C. Wyatt, "Automation bias: a systematic review of frequency, effect mediators, and mitigators", *Journal of the American Medical Informatics Association*, 19(1), 2012.
- ¹⁵⁹ J. Solomon, "Customization bias in decision support systems", *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 2014.
- ¹⁶⁰ For example, if a human has to rank a series of potential targets by priority when scripting a planning system, they will need to make value judgments about those targets that may be subject to harmful biases.
- ¹⁶¹ M.L. Cummings, "Automation and Accountability in Decision Support System Interface Design", *The Journal of Technology Studies*, 1 January 2006.
- ¹⁶² N. Renic and E. Schwarz, "Crimes of Dispassion: Autonomous Weapons and the Moral Challenge of Systematic Killing", *Ethics & International Affairs*, 37(3), 2023, pp. 321–343: <https://doi.org/10.1017/S0892679423000291>. In 2024, anonymous Israeli intelligence officers described how a targeting database called Lavender, which was intended to serve as a decision support tool, rather than an automated decision tool, "did it [targeting] more coldly." B. McKernan and H. Davies, "The machine did it coldly: Israel used AI to identify 37,000 Hamas targets," *The Guardian*, 3 April 2024. <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes>.

In some cases, organizations and decision makers might regard DSS simply as boxes that need to be checked.¹⁶³ This could undermine their capacity to properly vet and validate the system outputs or scrutinize their appropriateness for a given context. For example, a tool that indicates that the blast radius of a given weapon will not reach a civilian object could be treated as “proof” that due diligence was taken in the lead-up to the strike or as evidence to gain approval for a strike,¹⁶⁴ even if those outputs could not have accounted for dynamic factors in the environment that, if integrated, would have led to a different output. In cases where an action is contingent on a specific type of output from a DSS, operators might even tweak its inputs or parameters in order to generate that desired output,¹⁶⁵ rather than taking an initial undesired output or a relevant contextual factor as a signal to re-evaluate the objective or plan.¹⁶⁶

The Differences Between Civilian and Military Decision Support Systems

The growth of decision support in the civilian realm is often cited as evidence that similar technologies will see similar growth in the military domain. There are obvious parallels between, for example, social network analysis for counterinsurgency operations and analytics for digital marketing, or between resource optimization for multi-domain operations and optimization for ride-hailing apps.¹⁶⁷ Given these parallels, many have suggested that these same technologies could be leveraged to directly enhance computerized decision support in military operations.¹⁶⁸ However, there are important differences between the military domain and the civilian domain that could stand in the way of the transfer of these technologies from one to the other.

- 163 Interview with Lawrence Lewis, September 2021; J.R. Emery, “Probabilities towards death: bugsplat, algorithmic assassinations, and ethical due care”, *Critical Military Studies*, <https://doi.org/10.1080/23337486.2020.1809251>.
- 164 S.B. Sewall, *Chasing Success Air Force Efforts to Reduce Civilian Harm*, Air University, Air Force Research Institute, 2015, p. 158; J.R. Emery, “Probabilities towards death: bugsplat, algorithmic assassinations, and ethical due care”, *Critical Military Studies*, pp. 8 and 10: <https://doi.org/10.1080/23337486.2020.1809251>.
- 165 J. Rooke (interviewed January 2022) discussed how operators might, for example, craft an intelligence request that will be passed through a tasking DSS in such a way that improves the likelihood that it will be prioritized – in the same way that website owners will add certain keywords to their pages in order to improve their visibility in search engine results. Similarly, weaponeers performing a collateral damage estimation may adjust a strike plan so as to reduce the system’s estimate collateral damage – though they may do so in a good faith effort to reduce the risk of collateral harm, this might undermine consideration of whether such a strike should even be conducted in the first place. See: “CIVCAS Credibility Assessment Report (CCAR) for Allegation 1426, Ar Raqqa, Syria, 04 May 2017”, United States Central Command, December 2017, available for download at: <https://www.nytimes.com/interactive/2021/us/civilian-casualty-files.html>; “CIVCAS Credibility Assessment Report (CCAR) for 953, Tabqah Raqqah, Syria 01 March 2017”, United States Central Command, September 2017, available for download at: <https://www.nytimes.com/interactive/2021/us/civilian-casualty-files.html>.
- 166 “Al Hawijah ISIL VBIED Factory strike”, United States Central Command, 2015, p. 29/064, available for download at: <https://www.nytimes.com/interactive/2021/us/civilian-casualty-files.html>. According to anonymous Israeli intelligence officials, in the first weeks of the war in Gaza in 2023, a targeting algorithm’s parameters were tweaked so as to lower the threshold for the system to identify individuals as members of a combatant group. “‘The machine did it coldly’: Israel used AI to identify 37,000 Hamas targets,” *The Guardian*, 3 April 2024. <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes>.
- 167 When an app selects the optimal car for each driver based on factors such as each customer’s origin and destination relative to each car’s location and availability – it is solving an optimization problem that resembles that of a DSS that recommends a particular aircraft to strike a target in armed conflict. See: V.N. Gadepally *et al.*, “Recommender Systems for the Department of Defense and Intelligence Community”, *Lincoln Laboratory Journal*, Vol. 22, No. 1, 2016.
- 168 L.M. Zhang, “SAF exercise command post’s task made easier with AI and data analytics”, *The Straits Times*, 23 September 2021: <https://www.straitstimes.com/singapore/saf-exercise-command-posts-task-made-easier-with-ai-and-data-analytics>; CRS, “Joint All-Domain Command and Control (JADC2)”, CRS, Washington, 21 January 2022.

One crucial difference is that whereas many civilian application spaces are relatively linear – we know that there will be more demand for taxis during rush-hour or when it rains – military problem spaces are enormously non-linear.¹⁶⁹ In the military domain, there are also fewer tangible “ground truths” on which to anchor decisions; the “intelligence” that serves as the basis for decisions often relies on supposition to fill in the blanks. In a dynamic conflict, it is also harder to quantify success¹⁷⁰ or, by extension, code systems with an objective that reliably maximizes successful outcomes.

In the military domain, conditions will also be harsher, increasing the uncertainty in data. Adversarial actions such as subterfuge and hacking will undermine the integrity of data sources. Adversary forces will behave in unpredictable ways and will constantly modify their behaviour, thus rendering obsolete systems trained on data of their previously observed behaviours¹⁷¹ or systems relying on assumptions that held true in past operations.

Unlike the controlled environment of a digital game – an area where machine learning systems have shown significant performance in strategic planning – military domains will be marked by imperfect information, inconstant, ill-defined “rules” and shifting dynamics and adversaries. All of these factors may make it challenging to achieve consistently high performance in real life.¹⁷² There are also generally far fewer data available for developing military DSS, compared to the data available for civilian applications, meaning that these systems could be more subject to failure.¹⁷³

Most importantly, the safety-criticality of military applications, especially in the use of force, is much higher than it is for many of the most commonly cited civilian applications. System errors that would be unlikely to result in harm in the civilian domain (say, an app tasking a taxi that is not optimally placed for the user) would be unacceptable in a military context (for example, a planning tool tasking the wrong type of missile to strike a target).¹⁷⁴ This criticality not only makes the application of complex DSS to military roles more difficult than its application to civilian tasks, but also qualitatively different.

¹⁶⁹ S.J. Banks, “Lifting Off of the Digital Plateau With Military Decision Support Systems”, Master’s Thesis, School of Advanced Military Studies, United States Army Command and General Staff College, 2013, p. 28; and J.A.P. Smallegange *et al.*, *Big Data and Artificial Intelligence for Decision Making: Dutch Position Paper*, NATO Science & Technology Organization, Brussels, STO-MP-IST-160, 2018, p. 2.

¹⁷⁰ E. Wiseman, *Deep Learning for Human Decision Support*, Defence Research and Development Canada, Ottawa, 20 January 2017, p. 26; and V.N. Gadepally *et al.*, “Recommender Systems for the Department of Defense and Intelligence Community”, *Lincoln Laboratory Journal*, Vol. 22, No. 1, 2016, pp. 80–81.

¹⁷¹ A. Holland Michel, *Known Unknowns: Data Issues and Military Autonomous Systems*, United Nations Institute for Disarmament Research, Geneva, 2021.

¹⁷² M. Walsh *et al.*, *Exploring the Feasibility and Utility of Machine Learning-Assisted Command and Control, Volume 1, Findings and Recommendations*, RAND Corporation, Santa Monica, 2021, p. 63.

¹⁷³ Interview with Peter Svenmarck, November 2021; interview with Margarita Konaev, October 2021; interview with Svetlana Yanushkevich, November 2021.

¹⁷⁴ Interview with Maria Riveiro, November 2021; J. Dummon *et al.*, *Responsible AI Guidelines in Practice*, Department of Defense, Defense Innovation Unit, 2021, p. 16.

SECTION 5

IMPLICATIONS OF COMPLEX DECISION SUPPORT SYSTEMS FOR DECISION-MAKING IN THE USE OF FORCE

5.1 THE SHRINKING SPACE FOR HUMAN INTERVENTION

Though DSS do not “make” decisions on the use of force,¹⁷⁵ the sensor- and mathematics-based outputs of DSS in these roles have direct, serious humanitarian and legal implications.¹⁷⁶ Faithful implementation of international law calls for decisions on the use of force to be informed by contextual, value-based, human judgement.¹⁷⁷ When such decisions are supported by DSS, this judgement must include *contextual* assessment of the validity of the DSS output.

As described in Section 4, there are inherent challenges to the application of human judgement to DSS outputs. Any DSS output may be embedded with uncertainties, assumptions or biases whose existence or relevance to the decision, it might be claimed, cannot reasonably be known to the decision maker. Any unintended or unlawful harm arising from such unknowable issues might therefore be characterized as “blameless” – akin to the malfunction of a missile fuse or the jamming of a rifle – rather than the result of erroneous judgement, malign intent or failure of due diligence.¹⁷⁸ (see Section 5.2 below on accountability)

Each human judgement in a process leading to the use of force that draws on DSS outputs there-

¹⁷⁵ Only a human can make a decision, even if that decision is simply to do exactly what a output DSS proposes. For example, As noted in the course “An Introduction to the Collateral Damage Methodology (COM) and the Collateral Damage Estimate (CDE),” taught by the US Army Judge Advocate General’s School, Center for Law and Military Operations (CLAMO), the Collateral Damage Estimate for any given strike on any given target is “not itself a decision”.

¹⁷⁶ For example, if a DSS algorithmically ranks targets in a strike list according to factors such as their location, this has a direct and essential bearing on the eventual decision to attack or not attack any target on that list. K. Klonowska, “Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare”, Asser Research Paper 2021-02, p. 25; J.R. Emery, “Probabilities towards death: bugsplat, algorithmic assassinations, and ethical due care”, *Critical Military Studies*, p. 7: <https://doi.org/10.1080/23337486.2020.1809251>. Tools that indicate the location or attributes of civilian structures play a key role in supporting decisions on distinction, proportionality and precaution. M. 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, Art. 6, 2018, p. 80. An “imminent” threat detection alert by a predictive system could serve as the basis for the commander of a mission to claim the right to operate under less stringent collateral damage estimation requirements (this practice, sometimes referred to as self-defence rules of engagement, has reportedly led to numerous civilian casualty incidents in recent years). D. Philipps and E. Schmitt, “How the U.S. Hid an Airstrike That Killed Dozens of Civilians in Syria”, *The New York Times*, 13 November 2021: <https://www.nytimes.com/2021/11/13/us/us-airstrikes-civilian-deaths.html>.

¹⁷⁷ ICRC, *ICRC Position on Autonomous Weapon Systems*, ICRC, Geneva, 2021, p. 7; B. Wagner, “Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems”, *Policy & Internet, Special Issue: Internet Architecture and Human Rights*, Vol. 11, Issue 1, March 2019.

¹⁷⁸ Especially if operators are aware that they are unlikely to be held accountable for errors. This lack of accountability also likely makes operators less effective at identifying system errors. L.J. Skitka, K. Mosier and M.D. Burdick, “Accountability and automation bias”, *International Journal of Human-Computer Studies*, 52, 2000.

fore becomes a decision that could potentially be attributed to unanticipated technical malfunctions. Meanwhile, as DSS become more complex, the scale, complexity and unknowability of their uncertainties, assumptions and biases grow. Each new data source carries its own unique uncertainties and each new parameter is embedded with its own novel assumptions and biases.¹⁷⁹

Even when a DSS performs exceptionally well, if its uncertainties and assumptions and biases are not fully appreciated by the user, it is unclear whether its use would conform with faithful application of the law.¹⁸⁰ For example, consider a case in which a series of strikes that are carried out on the recommendation of a targeting DSS result in zero civilian casualties or damage to civilian objects. If nobody along the chain of command sufficiently understands the system and how it works, but instead trusts it on the basis of its performance alone, this positive record cannot on its own serve as proof that the humans carrying out the strikes have complied with their legal obligations under IHL.

Increased system complexity also blurs the line between decision support and the decision itself. In the case of a sufficiently complex nondeterministic system, the human “decision” that its output supports is reduced to a single binary judgement: whether to “trust” the system or not.¹⁸¹ The DSS becomes akin to a *decision-making system*, further eroding the human element in the application of force and further mirroring the concerns raised in regard to autonomous weapons.

The expanding use of DSS could also amplify the potentially detrimental effects of remoteness.¹⁸² Humans who are once or multiple times removed from the object of their decision could become detached from the dynamics and implications at play and at stake in their decision. This could lower their effectiveness¹⁸³ and lead them to take less care (and be more comfortable with unacceptable biases, assumptions and uncertainties) when making critical decisions on the use of force.¹⁸⁴ When testing or reviewing such systems, it could be difficult to assess whether users will develop a sense of remoteness or if they will retain “agency” in their decision-making. In this regard, DSS may raise some of the same challenges that have been identified with the application of rigorous human control and judgement over the use of force in the use of autonomous weapons¹⁸⁵ and remotely operated combat drones.

¹⁷⁹ M.L. Cummings, “Automation Bias in Intelligent Time Critical Decision Support Systems”, AIAA 1st Intelligent Systems Technical Conference, Chicago, September 2004; A. Naseem *et al.*, “Decision support system for optimum decision making process in threat evaluation and weapon assignment: Current status, challenges and future directions”, *Annual Reviews in Control*, 43, 2017, pp. 169–187.

¹⁸⁰ Understandability and transparency are regarded as being fundamental to the responsible use of AI. A.F.T. Winfield *et al.*, “IEEE P7001: A Proposed Standard on Transparency”, *Frontiers in Robotics and AI*, 26 July 2021.

¹⁸¹ Such “decisions” have been likened to simply pushing an “I-believe button” or exercising “rubber stamp” control. S.J. Freedberg Jr., “How AI Could Change The Art Of War”, *Breaking Defense*, 25 April 2019: <https://breakingdefense.com/2019/04/how-ai-could-change-the-art-of-war/>; R. Binns, “Human Judgment in algorithmic loops: Individual justice and automated decision-making”, *Regulation & Governance*, Vol. 16, Issue 1, January 2022.

¹⁸² K. McKendrick, *The Application of Artificial Intelligence in Operations Planning*, NATO Science & Technology Organization, Brussels, 2017, STO-MP-SAS-OCS-ORA-2017, p. 12.

¹⁸³ For example, one study discusses how ATR in three field exercises and in Bosnia actually tended to increase operator workload because they lost the capacities that they usually used for making decisions, such as “learning, history of past events, and surrounding contextual information.” M.A. O’Hair, B.D. Purvis and J. Brown, “Aided versus automatic target recognition”, *SPIE Proceedings Volume 3069, Automatic Target Recognition VII*, 1997.

¹⁸⁴ M.L. Cummings, “Automation and Accountability in Decision Support System Interface Design”, *The Journal of Technology Studies*, 1 January 2006. B. McKernan and H. Davies, “‘The machine did it coldly’: Israel used AI to identify 37,000 Hamas targets,” *The Guardian*, 3 April 2024. <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes>.

¹⁸⁵ ICRC, *ICRC Position on Autonomous Weapon Systems*, ICRC, Geneva, 2021, p. 7.

These challenges could be a particular concern in the use of DSS “at the edge”, for instance, in weapon sights or augmented reality headsets. Not only might operators at the edge have less time and cognitive capacity (given other demands upon their attention) to properly assess a given DSS output, but they may also have less access to broader contextual information. As a result, there will be fewer steps in the chain of human decision-making between that output and the use of force itself, as compared to faulty decisions that are made earlier in the process. This could also pose challenges with respect to highly automated systems. Though these systems may reduce the total cognitive load of operators, the complexity of these systems would result in a high cognitive load should they face a situation where they have to manually vet the system’s calibration or replicate its processes manually in order to certify the outputs.

As a result, the expanding use of more complex DSS, including those incorporating machine learning, in decisions on the use of force is likely to reduce and hinder the application of human judgement. Thus, it could significantly shrink the space for human intervention in the overall process.

5.2 ACCOUNTABILITY IN DECISION-MAKING

This shrinkage could put responsibility in the hands of a smaller number of decision makers along the chain of command,¹⁸⁶ potentially even to a disproportionate degree, such that individual DSS users are held fully responsible for failures that fall beyond their scope for legally mandated judgement.¹⁸⁷

In cases where a single DSS error could perpetuate across the entire chain of decision-making, it could diffuse responsibility among a large group of actors,¹⁸⁸ each of whom might skirt the claim that they could be fully responsible for the harm that resulted from a failure of contextual judgement.

Given that failures can be caused by technical as well as human-machine interaction problems, such failures might also implicate the developers and manufacturers of these DSS or the providers of data. It is widely observed that it would be challenging to attribute responsibility to actors so far removed from the use of force itself.¹⁸⁹

These factors mirror the potential “accountability gap” that could apply to the use of autonomous weapons.¹⁹⁰

¹⁸⁶ M. 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, Art. 6, 2018, p. 83.

¹⁸⁷ That is, by placing blame for harms on individual users of automated systems who become “moral crumple zones” for both the technical malfunctions of the system and the constellation of human and structural shortcomings the bridged these malfunctions into harm. M.C. Elish, “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction”, *Engaging Science, Technology, and Society*, Vol. 5, 2019: <https://doi.org/10.17351/ests2019.260>.

¹⁸⁸ J. Simpson, “Operations in deception: corrupting the sensing grid of the enemy”, *The Forge*, 2021.

¹⁸⁹ In some countries, companies may even be shielded from liability for such failures. B.A. Coleman and J.M. Moore, “Government Contractor Defense: Military and Non-Military Applications”, *American Bar Association Practice Points*, 12 September 2016: <https://www.americanbar.org/content/aba-cms-dotorg/en/groups/litigation/committees/products-liability/practice/2016/gvt-contractor-defense-military-non-military-applications/>.

¹⁹⁰ T. Chengeta, “Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law”, *Denver Journal of International Law & Policy*, Vol. 45, No. 1, Fall, 2016. R. Crootof, “War Torts: Accountability for Autonomous Weapons”, *University of Pennsylvania Law Review*, Vol. 164.

5.3 UNPREDICTABILITY, ERRORS AND CYBER VULNERABILITIES

As the number and breadth of tasks that systems are used for grow and the algorithmic architectures by which they operate become more complex, including through the use of machine learning, it can also become more challenging to test a system for all potential sources of errors or to identify when a system is encountering a context for which it was not specifically designed or validated,¹⁹¹ leading to “known unknown” failures that can be vexing for traditional channels of accountability in war.¹⁹² Machine learning-based systems, in particular, expand the scope for potentially unaccountable harm, given their heightened susceptibility to error modes that are impossible to anticipate in testing, prior use and reviews,¹⁹³ and also given their higher propensity to fail in unpredictable ways.¹⁹⁴

Introducing DSS in the use of force also expands the possibility of failures arising from cyber-attacks, which further confounds efforts to ensure effective human judgement in decisions on the use of force and accountability for such decisions. Every new computerized system deployed in support of decisions in the process leading to the use of force is a potential target for cyber attacks.¹⁹⁵ Successful hacking attacks on DSS could lead to unintended (and potentially grave) harm that cannot necessarily be attributed to a particular person or stakeholder. This risk, like others, could be elevated in machine learning-based DSS systems, which can be vulnerable to a range of types of adversarial techniques that cause the system to generate unpredictable erroneous outputs, often in a manner that is undetectable to human operators.¹⁹⁶

These challenges could be multiplied when numerous DSS contribute to decisions in a single process leading to the use of force, as any individual decision would implicate DSS that the individual decision maker may not directly interact with. This could be especially problematic if these individual outputs are automatically linked. Consider, for example, a so called “kill-chain” in which one DSS identifies a “threat”, passing that output to a second DSS that develops a plan to engage the threat, on to a third that devises a weaponizing option.¹⁹⁷ Not only could an error in any of these linked DSS cascade easily across the cycle, but there would be fewer humans to act as decision gateways through which such DSS outputs would have to pass. Furthermore, those humans would likely lack the capacity to properly evaluate every facet of such linked outputs. In this regard, the use of DSS could pose many of the core humanitarian concerns and legal challenges raised by the use of unpredictable autonomous weapons.

¹⁹¹ Interview with Peter Svenmarck, November 2021; M.A. Flournoy, A. Haines and G. Chefitz, *Building Trust through Testing: Adapting DOD's Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems*, WestExec Advisors, Washington, 2020; M. Luckcuck et al., “Formal Specification and Verification of Autonomous Robotic Systems: A Survey”, *ACM Computing Surveys*, Vol. 52, No. 5, September 2005.

¹⁹² It could be difficult to claim that harms arising from decisions (including decisions far removed from the actual application of force) made on the basis of DSS exhibiting such wholly unpredictable failures are the responsibility of the humans who made this decision in good faith, unaware that the failure had arisen. A. Holland Michel, *Known Unknowns: Data Issues and Military Autonomous Systems*, United Nations Institute for Disarmament Research, Geneva, 2021.

¹⁹³ K. Klonowska, “Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare”, *Asser Research Paper* 2021-02, pp. 9–14.

¹⁹⁴ A. Holland Michel, *Known Unknowns: Data Issues and Military Autonomous Systems*, United Nations Institute for Disarmament Research, Geneva, 2021.

¹⁹⁵ Interview with Svetlana Yanushkevich, November 2021; interview with Peter Svenmarck, November 2021; K. McKendrick, *The Application of Artificial Intelligence in Operations Planning*, NATO Science & Technology Organization, Brussels, 2017, STO-MP-SAS-OCS-ORA-2017, p. 9; for a concrete case, see: C. Cimpanu, “Two Android apps used in combat by US troops contained severe vulnerabilities”, *ZDNet*, 19 December 2018.

¹⁹⁶ P. Svenmarck et al., *Possibilities and Challenges for Artificial Intelligence in Military Applications*, NATO Science & Technology Organization, STO-MP-IST-160-S1-5P, Brussels, 2018, pp. 2 and 7; M. Comiter, *Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It*, Belfer Center for Science and International Affairs, Harvard Kennedy School, Cambridge, 2019; B.J. Schachter, *Automatic Target Recognition*, Vol. 4, SPIE, Bellingham, 2020, pp. 298–302.

¹⁹⁷ A. Deeks, N. Lubell and D. Murray, “Machine Learning, Artificial Intelligence, and the Use of Force by States”, *Journal of National Security Law & Policy*, Vol 10:1, p. 10.

ANNEX

ROLES OF DECISION SUPPORT SYSTEMS IN THE USE OF FORCE

This section describes some of the common military decision-making processes relevant to the use of force in which DSS (see page 20) may play a role.

Intelligence Retrieval, Organization and Analysis is the process of accessing relevant information and drawing actionable conclusions from that information. DSS may be used in this process to search for, synthesize and, in some cases, visualize information in such a way to make it easily understandable and accessible.¹⁹⁸ For example, a search tool might enable a human to find every intelligence report that mentions a particular individual by name or every data clip from a given location in a given time window. Or a synthesis tool may summarize the findings of numerous intelligence reports.¹⁹⁹ Similarly, a tool might sort and prioritize retrieved information in a manner that resembles the function of a web search system.²⁰⁰

Intelligence Fusion refers to the process of correlating information from disparate data sources to identify relevant objects, features, patterns or other findings.²⁰¹ This is a crucial element of intelligence analysis because often data from the battlefield are only meaningful when correlated with relevant data from other sources.²⁰² DSS may serve in fusion roles by presenting intelligence that can then be fused manually or by computationally correlating data points in order to generate a single fused output.

Intelligence Preparation of the Battlespace – also known as Intelligence Preparation of the Operational Environment – is the process of identifying all features of an area that may be relevant to military action so that any decision can take these features and factors into account.²⁰³

¹⁹⁸ J. Schubert *et al.*, *Data Farming Decision Support for Operation Planning*, NATO Science & Technology Organization, Brussels, 2017, STO-MP-SAS-OCS-ORA-2017, pp. 3-12-3-13.

¹⁹⁹ Prioritization tools might, for example, prioritize tactical messages among military operators, in the same way that a smart email inbox might organize emails according to how urgent they are. See: R.E. Marmelstein, "TIPS – A system for contextual prioritization of tactical messages", *Proceedings of the 14th International Command and Control Research and Technology Symposium (ICCRTS)*.

²⁰⁰ See: D. Yoo *et al.*, "Intelligent Army Tactical Command Information System based on National Defense Ontology", *Journal of The Korea Society of Computer and Information*, Vol. 18, No. 3, March 2013, pp. 81-82.

²⁰¹ B. Connable, *Military Intelligence Fusion for Complex Operations: A New Paradigm*, RAND Corporation, Santa Monica, 2012.

²⁰² For example, in order to track a single object over long distances or extended periods without interruption, it may be necessary to fuse the various surveillance feeds through which it passes. Similarly, if an object on the battlefield cannot be identified using one source of information (say, if a radar picks up an unfamiliar shape on a road), fusion with a second source (say, a camera) can aid in the characterizations of that shape to be a truck or a person or a tank.

²⁰³ R. L. Wolfel *et al.*, "It's in There: Rethinking(?) Intelligence Preparation of the Battlefield in Megacities/Dense Urban Areas", *Small Wars Journal*, 3 February 2016: <https://smallwarsjournal.com/jrnl/art/it%E2%80%99s-in-there-rethinking-intelligence-preparation-of-the-battlefield-in-megacitiesdense-urb>; for a list of advanced IPB systems and their capabilities, see: A. Bergeron Guyard, *Self-improving inference system to support the intelligence preparation of the battlefield: Requirements, state of the art, and prototypes*, Defence Research and Development Canada, Scientific Report DRDC-RDDC-2014-R136, Ottawa, December 2014; see also: C. Grindle *et al.*, "Automating Terrain Analysis: Algorithms for Intelligence Preparation of the Battlefield", *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 2004.

Such features can include enemy structures, geography, civilian structures, specially protected objects and weather conditions. DSS may support this process by consolidating or displaying all such features on a mapping system or, in more advanced instantiations, by identifying or characterizing such features through analytics.²⁰⁴ Systems might also complement these data by providing planning support, for example by indicating routes or areas that are impassable,²⁰⁵ indicating the shortest route between points²⁰⁶ or illuminating areas that are not visible to the enemy.²⁰⁷

Target/Threat Detection, Recognition and Tracking are processes that enable militaries to identify, characterize and follow objects in an area of operation, including (in a defensive context) incoming threats that may need to be intercepted or averted and (in an offensive context) potential targets for attack. DSS might “identify” such potential targets or threats by detecting technical signatures or indicators that match the characteristics of a known type of target or threat. This can also be achieved by a process of “anomaly detection,” whereby a system identifies any object or phenomenon that does not match a baseline of “normalcy.”²⁰⁸ In the case of moving targets, tracking enables militaries to maintain awareness of the object’s location and obtain greater information about its characteristics and its possible future actions.²⁰⁹ These systems may be used to facilitate intelligence analysis of data at the planning and support levels for targeting,²¹⁰ or in direct relation to weapons employment. Most air defense systems include automatic or aided target recognition and tracking tools, and such tools would be an essential component of autonomous weapon systems.²¹¹

Target Analysis and Target System Analysis are processes that study targets and networks of targets in order to identify, for example, their capabilities, features, vulnerabilities and, more broadly, their significance in the context of a mission’s objectives.²¹² They help assess whether and how these targets should or should not be attacked. Related processes include network analysis, which evaluates related entities (such as suspected members of an adversary’s armed forces) in order to characterize their identities, roles, capacities and vulnerabilities. For example, an analysis of phone calls between suspected combatants may indicate that one individual receives far more calls than any other, which may serve as evidence that they have a leadership or coordination role in the organization. DSS may serve in these roles by providing data analytics or predictive functions.

Course of Action (COA) Development and Evaluation is the process of developing and selecting a plan by which to achieve an objective. Courses of Action take a wide variety of forms, ranging from relatively simple plans (e.g. a plan for a single vehicle to approach and attack a single position) to complex multi-stage, multi-actor plans (e.g. a coordinated attack by air and ground forces against a variety of positions, as in the case of the first few hours of a land invasion).

²⁰⁴ N.M. de Reus, P.J.M. Kerbusch and M.P.D. Schadd, *Geospatial analysis for Machine Learning in Tactical Decision Support*, NATO Science & Technology Organization, Brussels, 2021, MP-MSG-184-08.

²⁰⁵ D. Yoo et al., “Intelligent Army Tactical Command Information System based on National Defense Ontology”, *Journal of The Korea Society of Computer and Information*, Vol. 18, No. 3, March 2013, pp. 81–82; D. Yoo, S. No and M. Ra, “A Practical Military Ontology Construction for the Intelligent Army Tactical Command Information System”, *International Journal of Computers Communications & Control*, 9 (1), pp. 93–100; S. Riley, “A Shared View of the Battlespace”, *C4ISR*, Vol. 5, No. 2, March 2006.

²⁰⁶ For example, the ArcGis mapping software suite.

²⁰⁷ For example, the U.S. Navy Sniper-RT and Sniper/Counter-sniper systems.

²⁰⁸ As in the case of a missile defense system, for example.

²⁰⁹ Anonymous interview with government employee, October 2021; X. T. Nguyen, “Threat Assessment in Tactical Airborne Environments”, *Proceedings of the Fifth International Conference on Information Fusion*, 2002, p. 1301.

²¹⁰ For example, Singapore Defence Science and Technology Agency (DSTA) Automatic Target Detection (ATD); Australia Analyst’s Detection Support System (ADSS); US DoD Project Maven.

²¹¹ ICRC, “Artificial intelligence and machine learning in armed conflict: A human-centred approach”, *IRRC*, No. 102 (913), Digital technologies and war, 2020, pp. 463–479.

²¹² M. 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, Art. 6, 2018, p. 70.

Even fairly routine COAs may involve hundreds of individual steps, which must be evaluated through a process that itself may involve hundreds of analytical actions.²¹³ In this role, DSS may propose potential COAs.²¹⁴ Or they may provide analytical tools that aid in the comparison of various COAs,²¹⁵ for example, by describing how long it would take to carry out a given COA, what resources it would consume and the likely loss of human life or property that it would incur.²¹⁶

Resource Optimization and Allocation refers to the process of calculating how to conduct a given action with the greatest efficiency and probability of success using the available resources.²¹⁷ This process, which is relevant across military planning – from logistics²¹⁸ and scheduling to “tasking” weapons for missions²¹⁹ – can apply both to planned and dynamic offensive operations, as well as defensive operations.²²⁰ Weapons Target Assignment, a specific form of optimization, is the process of assigning specific weapons to specific targets in a scenario where multiple targets exist (and generally where the time available for planning is limited). The goal of this process is to achieve the highest probability of destroying or neutralizing all targets with the greatest efficiency.²²¹ DSS can serve in a wide range of optimization roles, since they may often be based on well-defined mathematical problems.

Modelling and Simulation is the process of running a particular plan or scenario prior to exe-

- 213 K. McKendrick, *The Application of Artificial Intelligence in Operations Planning*, NATO Science & Technology Organization, Brussels, 2017, STO-MP-SAS-OCS-ORA-2017, p. 4.
- 214 E.C. Teppan *et al.*, “A Flexible Toolkit Supporting Knowledge-based Tactical Planning for Ground Forces”, 16th International Command and Control Research and Technology Symposium, Quebec City, 2011, pp. 5–13; see, for example, the U.S. Army Advanced Field Artillery Tactical Data System (AFATDS).
- 215 See: A. Tolk and D. Kunde, “Decision Support Systems – Technical Prerequisites and Military Requirements”, 2000 Command and Control Research and Technology Symposium, June 2000, Monterey, CA, United States. One widely used such system is the U.S. Army Course of Action Development and Evaluation Tool (CADET). See also Austria’s C2DSAS development program; the NATO Tools for Operational Planning Functional Area Services (TOPFAS) networked software; the U.S. CAESARII/COA system, which is used in wargaming; and the U.S. Army Battlespace Terrain Reasoning and Awareness – Battle Command (BTRA-BC).
- 216 L. Ground, A. Kott and R. Budd, “Coalition-based Planning of Military Operations: Adversarial Reasoning Algorithms in an Integrated Decision Aid”, 2016: <https://arxiv.org/abs/1601.06069>.
- 217 See, for example, Omnisys BRO; ABAM (Aircraft Beddown Allocation Module); Joint Assistant for Development and Execution (JADE); and the Theater Battle Management Core System (TBMCS). A.M. Mulvehill and J.A. Caroli, “JADE: A Tool for Rapid Crisis Action Planning”, Air Force Research Lab, Rome, 1999.
- 218 For a list of such tools used by the U.S. DoD, see: “Joint Publication 3–35 Deployment and Redeployment Operations”, U.S. Department of Defense, 10 January 2018.
- 219 See, for example, the SAIC PRISM scheduling system. “Is PRISM just a not-so-secret web tool?”, Electrospace.net, 23 June 2013: <https://www.electrospace.net/2013/06/is-prism-just-not-so-secret-web-tool.html>.
- 220 For example, deciding what weapons to deploy against an attacker or to dispatch for an offensive attack (see “Weapons Target Assignment” below) or determining how to place forces and weapons around a defended area so that they can defend it more effectively and efficiently in the case of an attack – H. Xiaofeng and S. Shifei, “Study on the Resource Allocation in Urban Defense Engineering with Intentional Threats”, *Systems Engineering Procedia*, Vol. 5, 2012; and T. Tanergüçlü *et al.*, “A decision support system for locating weapon and radar positions in stationary point air defence”, *Information Systems Frontiers*, Vol. 14, 2012. See also Singapore DSTA Target Look-Ahead (TLA) system, which recommends strike zones and estimates the time it would take them to reach them with each available weapon. “Striking Smarter and Faster”, Singapore Defence Science & Technology Agency: <https://web.archive.org/web/20200810083701/https://www.dsta.gov.sg/programme-centres/information-pc/striking-smarter-and-faster>.
- 221 See, for example, U.S. Attack Operations Decision Aid (AODA), D. Pedersen *et al.*, “Decision Support System Engineering for Time Critical Targeting,” MITRE Technical Paper, Bedford, 1999, p. 2; H. Kim and Y. Cho, “New Mathematical Model and Parallel Hybrid Genetic Algorithm for the Optimal Assignment of Strike packages to Targets”, *Journal of the Korea Institute of Military Science and Technology* (한국군사과학기술학회지), Vol. 20, Issue 4, 2017; K. Zhang *et al.*, “Efficient Decision Approaches for Asset-Based Dynamic Weapon Target Assignment by a Receding Horizon and Marginal Return Heuristic”, *Electronics*, 9, 1511, 2020; H. Naeem *et al.*, “A Novel Two-Stage Dynamic Decision Support based Optimal Threat Evaluation and Defensive Resource Scheduling Algorithm for Multi Air-borne threats”, 2009: <https://arxiv.org/abs/0906.5038>.

cutting the action in order to determine its likely outcomes and understand the likely effects of one's potential actions. This is similar to how an advanced chess player will play out the game several moves ahead in order to help her assess what move to make next.²²² For example, a simulation might indicate how the destruction of an airfield might affect an adversary force's capacity to launch aerial operations or how the killing of a specific military leader could alter the overall command structure and effectiveness of an adversary force. Modeling and simulation can be used at every level of command for decision support.²²³ DSS may support these functions by providing computerized simulation or predictive capabilities.

Weapons Effect Modelling is the process of predicting the specific destructive effects of a type of weapon against a type of target and its contents and surroundings, based on factors such as the weapon's effective blast radius, hazard area, its explosive force, the characteristics of the target (for example, whether it is a person, a group of persons, a vehicle or an armored vehicle, or – if it is a building – the size and construction of the structure) and the target's environment (geographic features, density and characteristics of other people and objects nearby, etc.).²²⁴ A related role is the "probability of kill" calculation, which is used to estimate the likelihood that a given weapon will hit, kill, destroy, incapacitate or neutralize a target person or object. DSS can support weapons effect modelling by retrieving information and by executing modelling calculations.

Collateral Damage Estimation (CDE), which draws on weapons effect modelling, is the process of calculating the likelihood that an attack on a given target using a given weapon will result in civilian casualties or damage to civilian objects. According to the procedures applied by many modern militaries, if the estimated collateral damage for a given strike exceeds a certain threshold, decision makers must seek approval higher in the chain of command before carrying out the strike. A DSS can be used to generate a collateral damage estimate. Users input factors such as the type of weapon, its method of employment, the type of target and the presence and proximity of civilians or civilian objects, and the system will generate a CDE "score."²²⁵

Weaponneering is the process of selecting weapons for a specific attack and calculating the required quantity and method of delivery in order to achieve the goal of destroying, killing, neutralizing or incapacitating that target, while complying with procedural constraints and legal requirements, including for the protection of civilians.²²⁶ DSS can be used in the process to retrieve information, calculate the weapon's effect and optimize for a given objective.

²²² D. Wilton, "The Application Of Simulation Technology To Military Command And Control Decision Support"; D.T. Maxwell, "An Overview of the Joint Warfare System (JWARS)", MITRE Corporation Technical Paper, August 2000.

²²³ Interview with Herman le Roux, November 2021; J. Hanna *et al.*, "Course of Action Simulation Analysis", 10th International Command and Control Research and Technology Symposium, June 2005.

²²⁴ Militaries may maintain a library of mathematical models for predicting the effects that weapons in their arsenal might have against specific targets or types of targets. These models are based on factors including the size, shape and other physical properties of the target. See: *USAF Intelligence Targeting Guide*, U.S. Air Force, Pamphlet 14-210, February 1998. For another example, see: *US Ship Weaponneering and Estimation Tool* (SWET).

²²⁵ *Avoiding and Minimizing Collateral Damage in EU-led Military Operations Concept*, European External Action Service (EEAS), Brussels, February 2016. S.B. Sewall, *Chasing Success Air Force Efforts to Reduce Civilian Harm*, Air University, Air Force Research Institute, 2015, p. 158; anonymous interview with an NGO employee, September 2021; "An Introduction to the Collateral Damage Methodology (COM) and the Collateral Damage Estimate (CDE)", course taught by the US Army Judge Advocate General's School, Center for Law and Military Operations (CLAMO); deployed CDE tools include the US DCiDE system and the US FAST-CD system.




²²⁶ M. 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, Art. 6, 2018, pp. 7-8; *USAF Intelligence Targeting Guide*, U.S. Air Force, Pamphlet 14-210, February 1998; JTCG/ME Weaponneering System (JWS) is the main weaponneering toolkit across the U.S. military services. The system includes "mathematical models, which enable weaponneers to predict the effectiveness of weapons against most significant targets. Inputs to these methodologies include factors such as target characteristics (size, shape, and hardness) and delivery parameters (altitude, speed, dive angle, etc.). *Joint Targeting School Student Guide*, Joint Targeting Guide, Dam Neck, March 2017, p. 139.

Combat Assessment, which some militaries subdivide into **Battle Damage Assessment**, **Collateral Damage Assessment** and **Munition Effectiveness Methodology**,²²⁷ is the process of evaluating the effects of the use of force after the weapon has been used or the action has been carried out. Such an assessment looks to estimate whether the goal of the attack was achieved (for example, by certifying that a targeted human is dead or incapacitated or that a targeted vehicle or building is destroyed or neutralized) as well as to ascertain whether any unintended effects were generated – notably whether any civilians were harmed and/or civilian objects damaged. Such assessments are crucial for informing decisions on how to proceed, including the decision to re-attack the target or to launch an investigation into whether applicable laws have been violated.

²²⁷ See, for instance: U.S. Chairman of the Joint Chiefs of Staff Instruction, *Methodology for Combat Assessment*, CJCSI 3162.02, 8 March 2019.

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



ICRC

International Committee of the Red Cross
19, avenue de la Paix
1202 Geneva, Switzerland
T +41 22 734 60 01
shop.icrc.org
© ICRC, April 2024