HUMANITARIAN LAW & POLICY

Artificial intelligence in military decision-making: supporting humans, not replacing them

August 29, 2024, Accountability / Analysis / Artificial intelligence in military decision-making / Generating Respect for IHL / IHL / New Technologies / Special Themes / Type of post

13 mins read

Wen Zhou
Legal Adviser, ICRC

Anna Rosalie Greipl
Researcher, The Academy of International Humanitarian Law and Human Rights (Geneva Academy)

The desire to develop technological solutions to help militaries in their decision-making processes is not new. However, more recently, we have witnessed militaries incorporating increasingly complex forms of artificial intelligence-based decision support systems (AI DSS) in their decision-making process, including decisions on the use of force. The novelty of this development is that the process by which these AI DSS function challenges the human's ability to exercise judgement in military decision-making processes. This potential erosion of human judgement raises several legal, humanitarian and ethical challenges and risks, especially in relation to military decisions that have a significant impact on people's lives, their dignity, and their communities. It is in light of this development that we must urgently and in earnest discuss how these systems are used and their impact on people affected by armed conflict.

With this post, Wen Zhou, Legal Adviser with the International Committee of the Red Cross (ICRC), and Anna Rosalie Greipl, Researcher at the Geneva Academy of International Humanitarian Law and Human Rights, launch a new series on artificial intelligence (AI) in military decision-making. To start the discussion, they outline some of the challenges and risks, as well as the potential, that pertain to the use of AI DSS in preserving human judgement in legal determinations on the use of force. They also propose some measures and constraints regarding the design and use of AI DSS in these decision-making processes that can inform current and future debates on military AI governance, in order to ensure compliance with international humanitarian law (IHL) and support mitigating the risk of harm to people affected by those decisions.

The urge to create technological solutions that can assist humans in decision-making is not recent. Yet, as the development of DSS advances, especially through the integration of powerful AI techniques, such as machine learning, we witness how various technologically advanced military powers started to pursue and utilize this military capability across a *wide range of tasks and many different levels of command*: from recommendations on military strategy to logistical decisions concerning the deployment of personnel or the transport of weapons.

More recently, machine-learning (ML)-based DSS supporting decision on the use of force are becoming an increasingly prominent military application of AI. *These computerized tools* provide complex assessment and nuanced outputs (i.e. analyses, recommendations and predictions) to aid humans in making complex decisions, such as who or what to attack and where, when and how. Using predetermined features and self-learned patterns, these types of AI DSS can highlight certain data points and prioritize and select features without human intervention. Hence, they are likely to have a significant impact on decisions to use force, including the process by which such decisions are made and the consequences that may result.

AI DSS is often *cited by militaries* as a key enabling technology *for faster and enhanced decision-making cycles*, touted as presenting critical military benefits and facilitating human decision-makers ability to comply with IHL and minimizing the risks for civilians.

However, careful consideration must be given to how these AI DSS operate, how humans interact with their outputs and what tasks they are suited for, as the human cost of resorting to these technologies will be heavily influenced by those factors. Importantly, AI DSS cannot improve targeting methodologies and policies that do not comply with IHL. Implementing AI DSS within such frameworks may only exacerbate unlawful or harmful effects, replicating and deploying them more quickly and widely.

Preserving human judgement in IHL

Under IHL, it is parties to armed conflict – *ultimately, human beings* – who are responsible for complying with IHL and who must be held accountable for violations. Hence, it is humans who must exercise judgement in legal determinations, such as about the lawfulness of an attack. Of course, this does not mean that humans responsible for planning, deciding upon, or executing attacks cannot use technological support or computerized analyses, such as AI DSS, to inform their decisions on the use of force. For instance, as is common today, militaries might consider outputs from AI DSS to support *identification of individuals and objects, weaponeering, or collateral damage estimations*. The upshot: in legally mandated decisions – like whether an object or person can be lawfully targeted – these systems' outputs can *inform* but must not *replace* human judgement.

AI DSS in military decisions: a risk or a potential for facilitating IHL compliance?

The incorporation of AI into DSS for military decision-making on the use of force in armed conflicts adds a new dimension to existing challenges with non-AI-based DSS.

In fact, applying AI DSS – particularly ML-based DSS – can negatively impact human decision-making for various reasons. Evidence suggests that using these systems increases the likelihood of unforeseen errors, and perpetuates and propagates problematic *biases*, particularly against individuals or groups of a certain age, gender or ethnicity, or persons with disabilities. A single AI DSS error can compound or cascade across a planning and decision-making process when many AI DSSs build upon and contribute to military decision in a single process.

The use of ML-based DSS makes it also more difficult for human users to identify erroneous outputs due to the increasing difficulty for humans to understand *how and why* a system generates a certain analysis, recommendation or prediction. A trend that is amplified by humans' propensity to over-reliance on AI DSS outputs (so-called automation bias), which is *particularly acute in situations of high pressure* – such as military decisions on the use of force – where decision-makers are left with limited time to assess, cross-check, or challenge AI DSS outputs. The growing tempo of military operations and the drive to use AI DSS for faster decision-making are *contributing to this issue*. The real-world implication of this might be that someone plans, decides upon or launches an attack based on an AI-decision-support system's output, rather than actually assessing the attack's lawfulness – essentially acting as a *human rubber stamp*.

Furthermore, the technical features of AI DSS make their use not suitable to all tasks and context. AI performs well when given clearly defined goals that can be translated into mathematical problems, such as deciding what type of weapon to deploy or where to position forces and weapons. However, the contextual and qualitative evaluations required by IHL are unlikely to offer clear goals for an AI decision-support system since they are extremely complex and *cannot be converted* to mathematical formulas or numerical values. The availability of high-quality and representative training data is another indispensable factor to ensure the adequacy and utility of the use of AI DSS. Yet, armed conflicts are characterized by uncertainty and volatility, compounded by adversaries seeking to deceive one another, all of which *makes it hard to source representative, transferable data*.

As a result of these, the use of AI DSS has the potential to hamper the human judgement involved in military decision-making on the use of force in armed conflicts, thus raising new humanitarian, legal and ethical questions. This is why, in order to ensure that an AI DSS *supports* rather than *hinders* military decision-making – and assists in ensuring IHL compliance, parties to conflict must carefully assess its suitability for the specific task and context.

A critical measure is to include AI-DSS in the legal review of weapons systems that incorporate AI-DSS tools or where such tools are part of the ways in which a weapon system is designed or expected to be used. More broadly, states should consider conducting legal reviews of all AI DSS that form part of military decisions to use force in armed conflict, at a minimum as a matter of policy or good practice. In doing so, militaries must account for the technical features of AI DSS, the tendencies of human users interacting with them, and mounting pressure on military decision-makers to expedite the speed of their decision-making processes.

At the same time, *careful* and *responsible* use of AI DSS may facilitate quicker and more comprehensive information analysis, which can support decisions in a way that enhances IHL compliance and minimizes risks for civilians. For instance, the use of AI DSS can assist in collecting and synthesizing information from open-source repositories online about the *presence of civilians and civilian objects* or support weaponeering by recommending means and methods of attack that can best avoid, or at least minimize, *incidental civilian harm*. The efficiency of any such tool will depend on access to good quality data. However, it is not clear whether militaries are sufficiently investing in the building and maintenance of required datasets, in particular when it comes to ML-based DSS.

But beyond AI DSS' potential to support civilian harm mitigation as part of targeting decision-making cycles, militaries should also consider prioritizing investment in the development and use of AI DSS specifically for the purpose of enhancing the protection of civilians – which is largely a lacuna at the moment. This could consist in AI DSS supporting *specialized engineers in advising military cells to map not only infrastructures' locations but* also their connectivity and in recognizing distinctive emblems and signals that indicate protective status, to reduce potential reverberating effects attacks can have on civilians and their livelihoods. Ultimately, such measures would be crucial in enhancing compliance with the IHL obligations to *take constant care to spare the civilian population* and to take *all feasible precautions in attack*.

Ways forward

In light of these risks and challenges, further measures and constraints on the use of AI DSS in military decisions may be necessary, particularly in situations where there is a possibility of harm or death to people or damage to objects and more generally where the decisions are governed by IHL.

Our recently published reports (an *expert consultation report* and an *ICRC-commissioned consultant report*) propose some of these measures and constraints. Notably, preserving human judgement in military decision-making over the use of force in armed conflicts may require new approaches to existing challenges arising from the interaction between humans and AI DSS, as well as technical requirements relating to the use of AI DSS. Some approaches may help to address existing technical challenges (e.g. predictability, understandability, and bias), while others may help to improve human decision makers' ability to critically engage with and use AI DSS outputs (e.g. mitigating automation bias).

For instance, the design and use of AI DSS need to allow human users to comprehend and question the systems' output. This could be done by enabling human users to cross-check AI DSS outputs against data from another source, and ensure users retain adequate time and space to critically engage with the system outputs to uphold human judgement and deliberation in decisions on the conduct of hostilities. Certain constraints may also be required, such as limiting the usage of AI DSS to certain clearly defined tasks and/or to certain contexts.

Furthermore, in light of the growing military interest in integrating ML in DSS, particular caution is warranted when it comes to the use of these systems for prescriptive tasks (such as recommending courses of action) or make predictions (such as predicting the risk posed by the actions of certain individuals or groups). Depending on the training data informing the AI DSS output, the lawfulness of using such recommendations or predictions in military decision-making on the use of force *may be called into question*.

Some usage of AI DSS may also be ruled out altogether. One clear example would be that such tools *must never be incorporated in nuclear-weapon command-and-control systems*. Constraining the use of AI DSS with continuous learning functions may also be necessary due to their particularly unpredictable nature.

We hope these recommendations and considerations can inform debates at international and national levels about the application and possible further development of international law and policies governing military applications of AI in armed conflicts.

Multistakeholder discussions on international governance of military applications of AI have significantly evolved in recent years including notably at the Responsible Artificial Intelligence in the Military Domain (REAIM) Summit (the *inaugural conference* in February 2023 in The Hague, and the upcoming *second conference* in September 2024 in Seoul). That said, the use of AI and ML in weapon systems has been touched upon at the Convention on Certain Conventional Weapons Group of Governmental Experts on Lethal Autonomous Weapon Systems since a decade ago, and more recently in the UN General Assembly (*Resolution 78/241*). Some states and militaries have recently developed AI principles and policies to govern military applications, including the United States' *Political Declaration on the Responsible Military Use of AI and Autonomy*, China's *Global AI Governance Initiative*, the French Ministry of Defence's *Artificial Intelligence in Support of Defence* report, the United Kingdom's *Defense Artificial Intelligence Strategy*, the Japanese Ministry of Defense's *Basic policy on the promotion of use of artificial intelligence*, and NATO's *Artificial Intelligence Strategy*.

In our view, all these and any future efforts should contribute towards ensuring that AI DSS, including ML-based DSS, remain means that *help* and *support*, rather than *hinder* or *displace*, humans in military decision-making. *Preserving human judgement* in decisions that pose risks to the life and dignity of people affected by armed conflict is essential to upholding ethical values and ensuring respect for applicable laws, including IHL.

******

Against this backdrop, we are pleased to launch a new blog series on "Artificial intelligence in military decision-making", through which we wish to contribute to the growing attention to and discussion on the use of AI in military decision-making, in particular towards the better understanding of the military use and associated humanitarian and legal implications of AI/ML DSS used in targeting processes and decisions to use force.

Considering the multifaceted challenges, risks but also potential of the military use of these systems, it is imperative to include voices from legal, ethical, policy, and technical experts. Hence, this series gathers a diverse and interdisciplinary group of experts to present their thoughts and insights into the discussion on the opportunities, challenges, and risks that arise from the use of AI in military decision making.

The discussions and reflections of the myriad dangers and challenges posed by AI in military decision-making will certainly not end here, but will carry on over the next months and years. We hope readers will enjoy reading this series and be inspired to engage in this timely and critical reflection.

See also:

- Ingvild Bode, *Falling under the radar: the problem of algorithmic bias and military applications of AI*, March 14, 2024
- Ruben Stewart & Georgia Hinds, *Algorithms of war: The use of artificial intelligence in decision making in armed conflict*, October 24, 2023
- Fiona Terry & Fabien Dany, *Harnessing the power of artificial intelligence to uncover patterns of violence*, May 25, 2023
- Christopher Chen, *The future is now: artificial intelligence and anticipatory humanitarian action*, August 19, 2021

Tags: AI, artificial intelligence, autonomous weapons, ethics, IHL, international humanitarian law, modern warfare, new technologies

You may also be interested in:

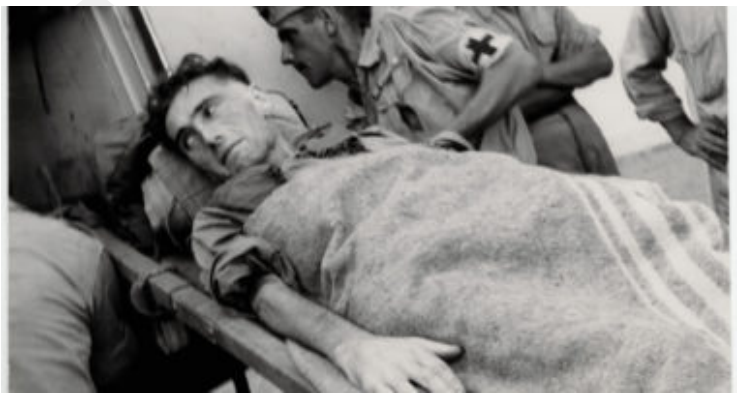
Beyond the rubble: eight overlooked ways that urban warfare is affecting children

14 mins read

Accountability / Analysis / Artificial intelligence in military decision-making / Generating Respect for IHL / IHL / New Technologies / Special Themes /

Type of post Timothy P. Williams, Alexandra Jackson & Vanessa Murphy

In cities from Gaza to those in Sudan and Ukraine, childhoods are irrevocably changed by ...



War's long legacy: the continued importance of the Geneva Conventions 75 years later

14 mins read

Accountability / Analysis / Artificial intelligence in military decision-making / Generating Respect for IHL / IHL / New Technologies / Special Themes /

Type of post Ellen Policinski

Next week marks the 75th anniversary of the four Geneva Conventions of 1949. Against a ...