Transcending weapon systems: the ethical challenges of AI in military decision support systems

September 24, 2024, Accountability / Analysis / Artificial intelligence in military decision-making / Conduct of Hostilities / IHL / New Technologies / Special Themes

🕐 12 mins read

Matthias Klaus
AI Ethics, Risk Analysis and International Security Studies expert

The military decision-making process is facing a challenge by the increasing number of interconnected sensors capturing information on the battlefield. The abundance of information offers advantages for operational planning – if it can be processed and acted upon rapidly. This is where AI-assisted decision-support systems (DSS) enter the picture. They are meant to empower military commanders to make faster and more informed decisions, thus accelerating and improving the decision-making process. Although they are just meant to assist – and not

replace – human decision-makers, they pose several ethical challenges which need to be addressed.

In this post, Matthias Klaus, who has a background in AI ethics, risk analysis and international security studies, explores the ethical challenges associated with a military AI application often overshadowed by the largely dominating concern about autonomous weapon systems (AWS). He highlights a number of ethical challenges associated specifically with DSS, which are often portrayed as bringing more objectivity, effectivity and efficiency to military decision-making. However, they could foster forms of bias, infringe upon human autonomy and dignity, and effectively undermine military moral responsibility by resulting in peer pressure and deskilling.

Autonomous weapon systems (AWS) steal the limelight in current military AI debates, but they are far from the only ethically challenging use cases. Here, I want to discuss AI-based decision support systems (DSS), which collect, aggregate and make sense of incoming intelligence, surveillance and reconnaissance (ISR) reports. Due to the proliferation of sensors, drones and the Internet of Things in military organizations, large amounts of data need to be processed at various military command levels. While this influx of data can result in better intelligence and coordination of own forces, its very existence poses challenges to analyzing and acting upon it in a timely manner. AI-based DSS are meant to cope with this issue and aid the command staff with *building common operational pictures, developing courses of action and supporting the execution of orders*, all within a fraction of the time needed by human planners.

However, while these systems are meant only to support and enable – rather than replace – human planners, they could stymie moral responsibility and foster unethical behavior and results instead. And while AI-based DSS have no embodiment *per se* and do not kinetically engage in warfighting themselves, they could have an even bigger impact than AWS. After all, AI-based DSS could shape the military decisions being made about the employment of AWS and human fighters alike. Thus, they deserve equal or even higher scrutiny, as their influence could be more subtle and impactful at the same time.

Expanding the ethical viewpoint

In the context of military operations influenced by AI, ethics play a special role. Ethics is concerned with how these technologies affect human values, human accountability and responsibility, and military virtues.

Military virtues such as courage, responsibility, and duty are fundamentally rooted in human judgment and decision-making. AI-based DSS, while valuable in processing large amounts of data, risk overshadowing these virtues. By taking on more of the cognitive load in military operations, AI-based DSS could dilute the human element of moral and ethical decision-making. A shift from human-led to AI-assisted judgment could erode the capacity of commanders to fully assume moral responsibility for their decisions, ultimately compromising military virtues.

The development and subsequent future deployment of AI-based DSS involves complex interactions between technology and human operators, going far beyond mere questions of functionality. Instead, it involves the evaluation of whether such systems can be aligned with fundamental principles such as human dignity and autonomy.

The ethical challenges associated with AI-based DSS are inherently linked with the technical aspects of these systems. For example, biases included in training data can lead to unwanted discriminatory outcomes, potentially violating the principles of justice and fairness. Opacity in AI systems can prevent humans from understanding or challenging the

systems' suggestions. This compromises the principles of transparency and accountability. These challenges affect operational efficiency, but also undermine the ethical principles at the core of military decision-making. This issue becomes even more pressing when it comes to maintaining meaningful human control over AI-based DSS, *which is often regarded as essential to ethical and responsible AI use in military operations.*

The rundown

The following examples represent a selection of ethical challenges related to AI-based DSS. They are deeply interconnected with concerns over human responsibility, military virtues and meaningful human control in military decision-making. These challenges affect the operational efficiency of decision-support systems but also raise critical ethical questions about the limitations of human decision-makers regarding their oversight, autonomy, and accountability in high-stakes environments.

Biases existing in the training data of AI systems can be unintentionally perpetuated or even amplified biases by the systems themselves. This can result in certain individuals or groups being discriminated against based on characteristics like sex, race, class, age, nationality, and many other factors. For *AI-based DSS, this could play a role* if its data is faulty, for example, if automated target recognition systems, which contribute to the creation of a common operational picture, misclassify certain individuals or groups as legitimate targets. There have been reports about this regarding drone warfare, where biased data labelling and interpretation have contributed to *people being targeted largely due to their apparent affiliation with certain tribes, for example.*

Explainability for AI-based DSS is a challenge, as most systems in development or research at this point are based on machine learning models, including *Convolutional Neural Networks* for image recognition, which are inherently opaque. This is problematic for an AI-based DSS, as it prevents its users from understanding why a certain course of action is proposed. What is more, it could also prevent human staff from identifying and correcting mistakes. Depending on the kind of system, methods and tools to explain the decision in question, and other factors, this could either be the military users or technical experts maintaining and training the systems, for example.

Automation bias refers to the *human tendency to over-rely on automated systems.* People often delegate tedious tasks to technology, *believing automated systems to have superior analytical abilities.* Automation bias manifests as errors of omission, where human operators miss anomalies that the system overlooks, or errors of commission, *where operators follow (faulty) suggestions without considering alternatives.*

AI-based DSS analyze data and suggest actions quickly and often more accurately than humans, leading to a natural trust in their recommendations. It could cause the users to disregard their training and intuition, relying on AI-based DSS outputs even when inappropriate. This is exacerbated if the system aligns with users' preferences, as they are less likely to question comfortable suggestions. Additionally, a lack of understanding of how these systems work can lead to *over-trusting the system*, especially if its limitations and biases are not apparent due to their opaqueness. Automation bias risks collateral damage and unnecessary destruction on the battlefield by causing operators to accept AI-based DSS' suggestions uncritically, potentially resulting in unnecessary suffering and harm.

Human autonomy faces risks with the use of AI-based DSSs in military operations. While automation bias affects planners, the execution of these plans by soldiers on the frontlines also poses significant challenges. AI-based DSSs could cultivate micromanagement, where individual soldiers receive detailed, granular orders via the system, potentially dictating routes, targets, and methods. This could erode human autonomy as fewer and fewer decisions are left to individuals. *Projects like extended reality visors for soldiers already hint at this future, displaying real-time information and targets.*

Comprehensive AI-based DSSs could foster a form of virtual remote command and control, reducing soldiers to executing orders displayed on their devices without critically engaging with these systems' outputs. This scenario risks soldiers not questioning orders, even if they have insights suggesting alternative actions. If soldiers receive commands about enemy positions via AI-based DSS, they might act without verifying the situation. This challenges the *military self-perception of conscious decision-making in the spirit of the "Auftragstaktik"* and in the worst case, result in soldiers "only following orders".

Deskilling refers to the *loss of professional skills due to lack of practice*, often resulting from technological advancements. In the military, the principle of "train as you fight" emphasizes the *importance of realistic training to maintain essential skills.* Soldiers must be proficient in both hard skills, like tactical behavior, and soft skills, such as decision-making and adaptability, to respond effectively in combat.

AI-based DSSs, while reducing the cognitive workload, risk deskilling command staff by taking over planning and decision-making tasks. Without regular practice, staff may lose proficiency in these, which can be crucial during system failures. For example, automated target recognition and threat assessments may lead to a degradation of "manually" assessing ISR reports. Ineffective human decision-making will expose troops, civilians, and enemies to unnecessary risks, leading to wrongful attacks, fratricide, or collateral damage. Additionally, *deskilling can erode military virtues* like courage and mercy, which require constant practice as well.

Acceleration pressure would be a direct result of the successful implementation of AI-based DSS in a military organization. *DSSs can significantly speed up the decision-making process,* but they risk setting an accelerated pace as the standard. Command staff may become resistant to slowing down the decision-making process to verify AI-based DSS results. Peer expectations for rapid results could make soldiers reluctant to voice concerns. As speed is a deciding factor on the battlefield, as echoed by the axiom *"a good plan violently executed now is better than a perfect plan next week"*, any form of slowing down to check results could result in peer pressure or bullying against cautious members of staff. Thus, AI-based DSS supporting this acceleration pressure hinders the *concept of meaningful human control.* In essence, it boils down to asking ourselves how we want to exercise control and how much of it we are willing to relinquish in favor of accelerating the decision-making process.

Human dignity in the context of AI-based DSS involves the ethical implications of AI calculating attrition rates for combat scenarios, estimating the likelihood of injuries or deaths among soldiers and non-combatants. While practical for military planning, this raises ethical concerns. Soldiers expect to risk their lives following human commanders who weigh the consequences, but they should not be reduced to statistics in an algorithm's cost-benefit analysis. This approach dehumanizes everybody affected by the calculation, including enemy troops and non-combatants affected by the decisions made.

While, for example, collateral damage estimation (CDE) is *already being performed with the aid of basic computer tools*, the extent to which AI-based DSS could change the process is cause for ethical concerns. It removes the decision-makers even further from the human element of warfare. Traditionally, these decisions are being made by human commanders bearing the moral responsibility for weighing the potential harm to their soldiers and civilians against the value of their objectives. In contrast, when such calculations are automated by AI-based DSS, the risk is that human lives could be reduced to mere data points in an algorithmic equation. This would obscure the moral significance of choices being made.

Looking ahead

The integration of AI-based DSS in military operations promises more efficient and faster decision-making. However, as illustrated, these systems also present significant ethical challenges that must be addressed to ensure their responsible use.

To mitigate these risks, it is crucial to develop robust ethical frameworks and guidelines for the use of AI-based DSS. Continuous training and education for military personnel on the limitations and potential biases of these systems are essential. This should include fostering critical thinking and a healthy portion of skepticism or caution towards DSS to allow responsible and meaningful human control.

The assumption that AI will inevitably change military operations, often referred to as "technological determinism", should be critically examined. Instead of simply accepting this notion, a comprehensive risk analysis is in order. This

would allow for a careful consideration of the ethical challenges raised earlier, including weighing potential risks against benefits. *Such an approach could indeed foster (responsible) innovation by providing guardrails developers and users could follow.*

It is essential to draw more attention to AI-based DSS. Even if they are meant to merely assist human decision-makers, they will inevitably shape their decisions by how they process data, present information and propose courses of action. By regarding AI-based DSS as *socio-technical systems*, we need to raise awareness to their moral impact. While they may lack moral agency, they are tools with moral impact to be deployed in armed conflicts. Therefore, we need to foster our discussion on how these systems should be designed, developed, used and overseen.

See also:

- Jimena Sofía Viveros Álvarez, *The risks and inefficacies of AI systems in military targeting support*, September 4, 2024
- Ingvild Bode, Ishmael Bhila, *The problem of algorithmic bias in AI-based military decision support systems*, September 3, 2024
- Wen Zhou, Anna Rosalie Greipl, *Artificial intelligence in military decision-making: supporting humans, not replacing them*, August 29, 2024

Tags: AI, AI decision support systems, algorithmic bias, artificial intelligence, autonomous weapon systems, AWS, conduct of hostilities, ethics, human dignity, IHL, LAWs, modern warfare
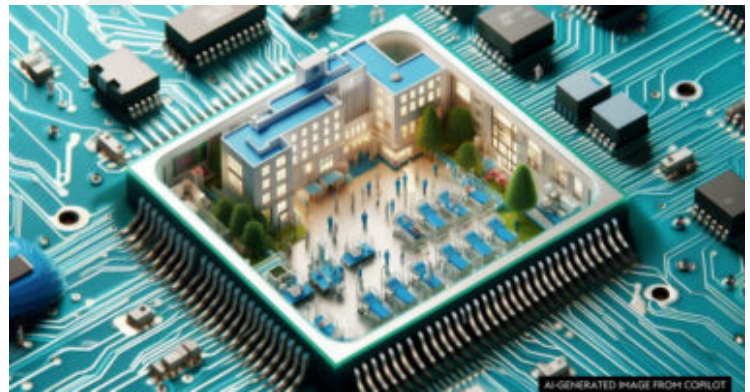
You may also be interested in:

Procedures for internment review under the Fourth Geneva Convention: reflections from New Zealand

🕐 15 mins read

Accountability / Analysis / Artificial intelligence in military decision-making / Conduct of Hostilities / IHL / New Technologies / Special Themes Tim Wood

Conceive, standardize, integrate: the past, present, and future of adopting distinctive emblems and signs under IHL

🕐 10 mins read

Accountability / Analysis / Artificial intelligence in military decision-making / Conduct of Hostilities / IHL / New

The Fourth Geneva Convention was the first humanitarian law convention dedicated to protections for civilians ...

Technologies / Special Themes Samit D'Cunha

When the very first Geneva Convention was adopted in 1864, it was the culmination of ...