



UNITED NATIONS
HUMAN RIGHTS
OFFICE OF THE HIGH COMMISSIONER

Photos by Unsplash: Ian Battaglia

BRIEFER ON **HUMAN RIGHTS AND ARTIFICIAL INTELLIGENCE IN THE MILITARY DOMAIN**

The design, development, deployment, use and regulation of artificial intelligence (AI) in the military domain has the potential to change all aspects of military affairs significantly. Military applications of AI can be used for the purposes of target identification, influence detention operations, and enable weapon systems to operate autonomously, including the application of force without human intervention.

While AI technologies in the military domain could be used to enhance protection for civilians, for example by leveraging AI to mitigate civilian harm, the military use of AI poses significant human rights and humanitarian concerns and raises legal, technical and operational challenges. It presents risks to the enjoyment of a wide range of human rights, including the right to life, the right to health, the right to food, the right to liberty, freedom from torture and ill-treatment and the right to be free from discrimination.

Importantly, the use of AI in the military domain is likely to reshape and redefine the roles and responsibilities of all actors involved, in particular States, their armed forces, and the business sector, in the design, development and deployment of these technologies. This requires not only a better understanding of the functioning of these technologies, but also clear answers and guidance on the human-machine interaction throughout the lifecycle of AI in order to ensure compliance with international law, in particular international human rights law and international humanitarian law.

ARTIFICIAL INTELLIGENCE IN THE MILITARY DOMAIN

Artificial intelligence is an umbrella term that covers machine-based systems and techniques that process data based on specific algorithms. It includes speech recognition, machine vision and language processing. In general, AI systems work by ingesting large amounts of labelled training data, analysing that data for correlations and patterns and using these patterns to make predictions.

Just as in the civilian domain, the possible applications of AI in the military are wide ranging and go beyond the battlefield. For example, AI can be used to automate certain tasks or enhance intelligence, surveillance and reconnaissance capabilities by processing vast amounts of data and providing recommendations.

AI has the potential to influence military decision-making at the strategic, operational and tactical levels and impact the overall speed and intensity of warfare.

AI IN WARFARE

AI technologies can be used to perform a variety of functions and tasks, ranging from:

- **AI and weapon systems:**

The incorporation of AI into weapon systems, including autonomous weapon systems can enhance their capabilities but also amplify their risks. A primary concern is related to the autonomy of such weapon systems, in particular with respect to the use of lethal force, allowing them to identify, select and engage targets and use lethal force with little or no human involvement. In addition, a weapon system enhanced by AI may possess adaptive capabilities enabling it to learn from its environment and adapt its behaviour throughout its deployment. This might significantly reduce the predictability of the system's behaviour potentially resulting in unintended actions. For example, a system might misidentify civilians moving in groups and carrying items resembling weapons as combatants, and as a consequence target them leading to civilian deaths.

- **AI and decision-support systems:**

AI decision-support systems (AI DSS) are tools that leverage AI to support decision-making processes. They rapidly collect, process and analyse large amounts of data, identify patterns and recommend courses of actions to decision-makers. AI technologies can, for example, assess real-time data from various sources, such as satellite imagery, drone feeds and sensors and provide commanders with a comprehensive view of the battlefield. Based on pre-defined criteria, AI-DSS can rapidly process surveillance and intelligence data, create profiles of people and places and subsequently identify targets. These systems have wide-ranging applications, influencing military actions such as targeting and detention, and can ultimately impact the overall speed and intensity of warfare. While AI-DSS are designed to support, rather than replace, human decision-making processes, humans tend to over-rely on recommendations provided by AI-DSS without fully understanding the technical aspects behind these outputs. There is a risk, as examples from the practice of armed forces show, that the nature of human machine interaction is happening in a way that limits the critical reflection space of human operators. Considering the limitations of AI,



including lack of transparency and biases, such overreliance during the execution of critical tasks could result in decisions that harm individuals and negatively impact human rights.

AUTONOMOUS WEAPON SYSTEMS

AI technologies could serve as a driver of autonomy in the execution of different functions, including the use of lethal force. There will, however, always be a human element and human responsibility whether at the phase of development or use. The human-machine interaction is becoming increasingly complex.

The UN Secretary General has called for the prohibition of lethal autonomous weapon systems that function without human control or oversight and which cannot be used in compliance with international law and to regulate all other types of AWS.

- **AI and cyber & information operations:**

AI is increasingly used in the cyberspace and information operations for both offensive and defensive capabilities. While AI can be employed to detect and respond to cyber threats in real time, it can also be used to execute cyberattacks. In information operations, AI can be used to change the way in which information is manipulated and can accelerate the creation and dissemination of content. Where this is used to spread and create disinformation, deepfakes, fake news and contradictory information, it could contribute to the erosion of trust in official sources, institutions and the people in general. Concretely, AI could be used to spread false information about evacuation routes or humanitarian aid, potentially causing physical harm and deaths or hindering access to essential resources like food, water and shelter. AI-driven information operations can thus have real-world consequences for individuals.

— HUMAN RIGHTS IMPLICATIONS OF AI IN THE MILITARY DOMAIN

- **Operational inaccuracies:**

Data-driven AI systems, such as machine learning inherently carry a risk of inaccuracy. It can be very difficult to verify their accuracy and identify the sources of errors. Much depends on the data used for the training of the model. For example, inaccurate data relating to the identity of individuals or the location of military objectives could lead to the misidentification of targets and unintended harm to civilians or civilian objects. Similarly, AI could amplify the negative consequences resulting from the erroneous identification of civilians and civilian objects as military objectives, potentially leading to a significant increase in civilians and civilian objects being wrongly struck. Such outcomes can have profound implications for the right to life, the right not to be subjected to cruel, inhuman or degrading treatment and a range of economic, social and cultural rights.

RECOMMENDATIONS

Verified and representative data are essential to lower the risk of unintended effects that could have negative human rights consequences. Rigorous testing and validation processes that involve diverse stakeholders can help identify and rectify biases. Cross-checking outputs can minimize negative effects.

While efforts to improve the quality of data is essential, the mitigation of operational inaccuracies ultimately depends on incorporating more measures to improving the processes and conditions for human-decision making and the role of AI in those processes. Mechanisms for continuous assessment and evaluation ensure that previous shortcomings are identified and addressed.

- **Discriminatory data:**

Algorithm-based decisions tend to reflect societal biases and can thereby reinforce or perpetuate existing discrimination due to the way in which datasets are curated. This is especially pertinent in the contexts of surveillance, detention and law enforcement operations. The systematic collection of personal data, including data revealing information about racial or ethnic origin, political opinions or religious beliefs, can expose individuals to discrimination if the data is not adequately reviewed for bias. Women, minority groups and marginalized people are often disproportionately impacted by prejudices and biases, which data-driven tools can perpetuate.

RECOMMENDATIONS

Adequate collection and review of data collection as well as review of algorithm design along with continuous monitoring for discriminatory outcomes is necessary to mitigate negative impacts on human rights. Ensuring the availability and use of quality datasets is a fundamental part of this process.

HUMAN AGENCY AND HUMAN RESPONSIBILITY

While AI can support military capabilities, the ultimate responsibility for its use and the consequences lies with humans. Machines cannot be held accountable for IHRL and IHL violations.

Humans retain the agency to decide if, when and how AI should be deployed and determine what data will be used.

A same technology is thus likely to have different human rights consequences depending on the use or user.

- **Lack of transparency:**

The lack of transparency around the design, development and deployment of AI systems can be an obstacle to understanding whether and how AI was used and based on what considerations the system generated a specific output. The attribution of actions to specific actors becomes difficult, making it challenging to establish responsibility and ultimately ensure accountability for violations of IHRL and IHL. In the same vein, the opacity of AI systems complicates the identification of potential biases. This issue is further compounded by the military's use of both open-source and classified data which restricts external oversight and hampers the development of effective safeguards, including to protect human rights.

RECOMMENDATIONS

Promoting the right to information through public access to data practices, sharing lessons learned from the development, deployment and use of AI in the military, along with best practices for testing, evaluation, and legal review processes, can contribute to a broader understanding of how these systems are developed and assessed. Independent and external auditing of AI systems should be facilitated.



- **Dual-use and repurposing of AI technologies:**

AI technologies, whether initially designed for civilian or military use, are inherently dual-use and readily repurposed and may blur the line between civilian use and military application. Civilian AI applications can be adapted for military activities, including in the context of cyber operations or weapons development, while military applications, such as facial recognition, are likely to be used in civilian contexts such as law enforcement and border control operations. This raises concerns around privacy, discriminatory outcomes, and lawful use of force. In particular, as there are differing standards on the use of force between conflict and non-conflict settings, the AI system must be adapted for its intended use, in order to ensure it operates in line with military or civilian requirements. Low-cost repurposing, combined with the widespread availability of these technologies, makes them accessible to both State and non-State actors, thereby increasing the risk of misuse and proliferation.

RECOMMENDATIONS

Actors repurposing or importing technologies across domains should test and adequately adapt them to ensure their lawful use in compliance with the applicable standards in the relevant context and domain.

- **Multi-dimensional nature of harm:**

Given the diverse applications of AI in military activities, the potential harm, identified in the previous bullets, that can result from the design, development, deployment and use of AI extends far beyond physical harm, such as death and injury. These effects can manifest in other forms, including psychological harm, the erosion of trust and violations of the right to privacy, the right to food and other abuses affecting people's well-being. Moreover, the harm done by an AI system should not be analysed in isolation as an AI system might not only cause harm but also exacerbate the harms due to other factors and systems, leading to cumulative effects, depending on the context in which the AI system operates.

RECOMMENDATIONS

Understanding the different types of harm that may be caused by AI is the first step in formulating effective and comprehensive recommendations, including mitigation measures, on the responsible development and military use of AI. This understanding is essential to mitigate potential harms effectively.

COMPLEMENTARITY BETWEEN IHL AND IHRL

While IHL provides rules on the conduct of hostilities, the means and methods of warfare and the treatment of protected persons in armed conflict, IHRL is applicable in peace and war time, including to the development, deployment and use of AI in the military domain. The complementary applicability of IHL and IHRL ensures a comprehensive legal framework to address the risks related to military AI applications regardless of the context in which they are used.



Photos by Unsplash: Luke Jones

— IMPORTANCE OF INTERNATIONAL HUMAN RIGHTS LAW

International human rights law provides a legally binding framework applicable throughout the entire lifecycle of AI, including its design, development, deployment, use and regulation. Under international human rights law, States are obliged to respect and protect individual rights, including in the face of new technologies such as AI. Human rights obligations and standards help clarify the roles and responsibilities of all actors involved and provide clear, unambiguous and universally recognized guidance on transparency, participation, non-discrimination and accountability, including access to justice and reparations for victims. Using a human rights framework to address the impact of AI can help leverage its positive aspects and mitigate and avoid harm.

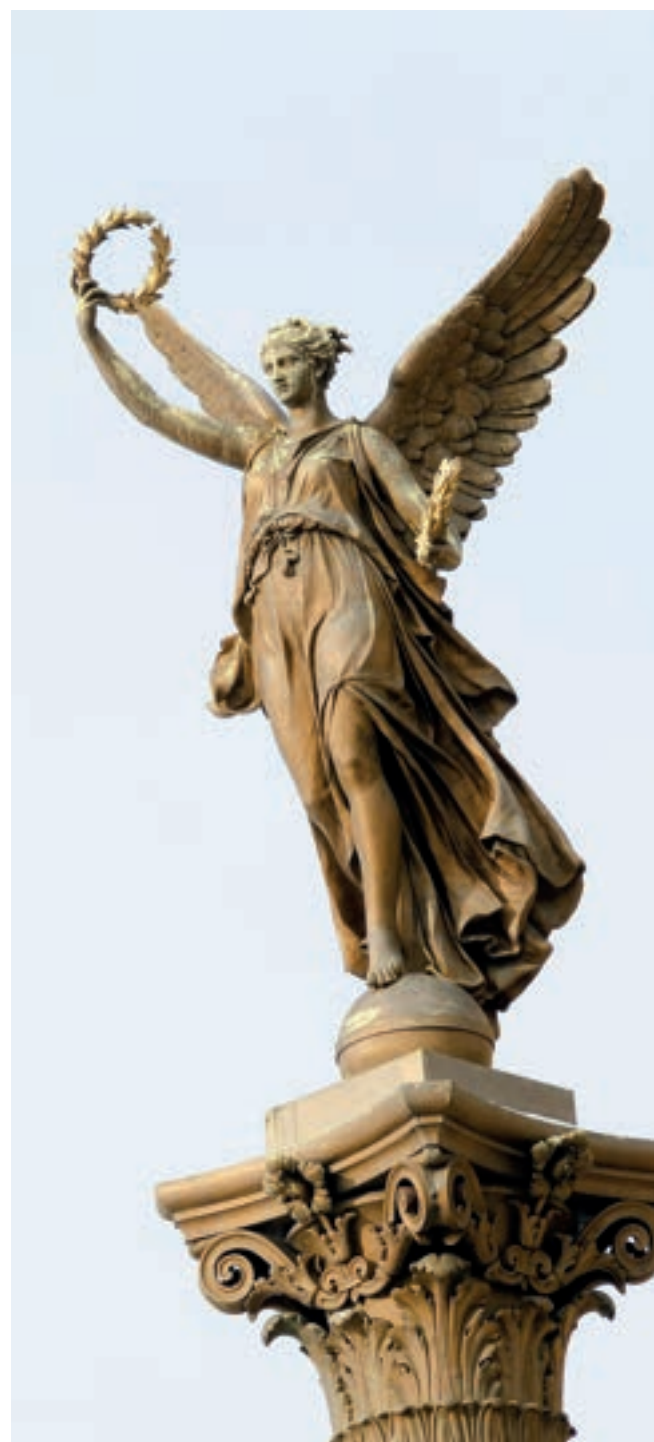
International human rights law promotes and provides clear and well-established guidance on:

Equality and non-discrimination – Integrating human rights helps ensure that new technologies, such as AI, do not discriminate or exacerbate existing inequalities. Disaggregated data can help understand the differentiated impacts of technologies, such as AI on various individuals and groups.

Transparency – States and companies should ensure transparency in the development and use of AI systems. They should make information available about the kind of systems they use, for what purposes and the safeguards in place to mitigate potential risks. Transparency also applies to making information available about the outcome of human rights impact assessments and ongoing reviews and oversight, as well as to sale and acquisition of AI systems. Transparency ensures that stakeholders understand both the role of AI in military operations and its impact on human rights, as well as its compliance with international law, and is necessary to ensure access to remedy and accountability.

Accountability – States must not only refrain from violating human rights but also take proactive steps to protect individuals from infringements by other actors, including the business sector. They must investigate and prosecute perpetrators of violations of IHRL and IHL and provide remedies, including reparations, to victims.

Inclusive and multistakeholder approaches to AI Governance in the Military Domain – Despite several national, regional and international initiatives, current discussions on AI governance in the military domain lack meaningful participation and are largely dominated by States and industry actors. An inclusive dialogue between governments, industry, military, civil society, and academia from all regions and States will help to address the legal and technical complexities of AI. This requires wider stakeholder engagement, including people from affected communities and their representative organizations, in discussions on AI governance that are anchored in human rights.



Photos by Unsplash: Luke Michael

RESPONSIBLE BUSINESS CONDUCT

Human rights norms and standards guide both States and businesses, requiring them to ensure that the development and the application of new technologies do not pose risks to the enjoyment of human rights. States should adopt and implement sectoral regulations and oversight mechanisms to prevent companies from circumventing their human rights responsibilities related to AI applications in the military domain. Importantly, States and businesses should ensure that human rights due diligence is conducted on an ongoing basis throughout the AI lifecycle, namely when AI technologies are developed, transferred, acquired and deployed. This includes an impact assessment to identify, mitigate

and prevent adverse impacts on human rights.

As part of human rights due diligence, once AI technologies are in use, States and companies should conduct regular review and oversight to identify, assess and mitigate human rights impacts across the full AI technology lifecycle, and make the findings of such reviews publicly available in the appropriate format, in line with the UNGPs. Human rights due diligence should meaningfully involve external experts and affected stakeholders. Where the use of AI technologies is found to have contributed to or directly caused human rights violations and abuses, States should discontinue their use until they are able to ensure that they are used in line with international human rights law, and contribute to remedial options.

BUSINESS AND HUMAN RIGHTS

AI innovation is largely driven by the private sector, with commercial advancements leading the military. The industry is led by a few global tech companies, influencing the defence acquisition process and providing technical services to the military. The private sector plays a key role in ensuring responsible AI.

The UN Guiding Principles on Business and Human Rights (UNGPs), which are grounded in international human rights law, provide a solid framework for responsible AI conduct by States and companies. They can help ground the design, development, deployment and use of AI in human rights. The OHCHR's B-Tech project seeks to improve the implementation of these principles in the technology sector, offering recommendations, tools and guidance on the application of the UNGPs, including addressing human rights risks posed by the AI technologies.

OHCHR SEEKS RESPONSIBLE MILITARY AI SYSTEMS THAT RESPECT IHRL AND IHL AND PROPOSES:

- Integration of IHRL through the life cycle of AI from design to development and use
- A ban on military AI systems that do not operate in compliance with IHRL and IHL
- A moratorium on sale, transfer and use of AI systems carrying high risk for human rights
- Use of IHRL to guide AI regulation and governance
- Use of rigorous data as a basis for training military AI systems

- Adoption of strong data protection legislation that protects personal data from unauthorized access, use and disclosure
- Monitoring of military AI systems by independent oversight bodies
- Use of human rights impact assessments in the development and use of military AI systems, including during armed conflicts
- Integration of affected individuals and communities, civil society organizations as well as States and businesses across regions in AI governance and regulation discussions
- Implementation of the UN Guiding Principles on Business and Human Rights to businesses involved in new and emerging technology in the military domain.



UNITED NATIONS
HUMAN RIGHTS
OFFICE OF THE HIGH COMMISSIONER

Palais des Nations,
CH-1211 Geneva 10, Switzerland
www.ohchr.org

