FORUM: DISARMEMENT COMMISSION
QUESTION OF: Measures to Establish Normative Frameworks Governing the Deployment of Artificial Intelligence Within Contemporary Warfare
MAIN SUBMITTER: Canada
CO-SUBMITTERS: France, Japan, Mexico, Indonesia, Brazil, Ethiopia, Croatia, Algeria, Australia, Colombia, Argentina, Ukraine, Nigeria, Bulgaria

DISARMEMENT COMMISSION,

Fully aware that existing International Humanitarian Law (IHL) applies to the conduct of hostilities involving artificial intelligence, while recognising the need for further clarification regarding its application to emerging military AI technologies,

Recognising the potential security and diplomatic sensitivities associated with international transparency initiatives, particularly concerning the disclosure of classified military information and national security interests,

Alarmed by the potential consequences of insufficient human oversight in the deployment and operation of autonomous military systems, particularly those capable of independently selecting and engaging targets,

Concerned by the potential for increasing automation in warfare to create psychological and moral distance between human decision-makers and the consequences of lethal force,

Acknowledging the importance of international cooperation in establishing common safety standards for the responsible development, testing and deployment of military AI technologies,

Aware that it becomes difficult to determine responsibility when an AI-enabled weapon contributes to an unlawful attack,

Deeply concerned by the possibility of civilian harm caused by AI/autonomous weapons,

Further aware that France has already developed ethical guidance for military AI, addressing responsible use and human decision-making.

Recognizing the rapid development of artificial intelligence and its application in contemporary military operations,

Concerned by the potential risks associated with AI-enabled weapon systems, including civilian harm, misidentification, and the loss of meaningful human control,

Recognizing the current ongoing discussions between the Convention on Certain Conventional Weapons and its Group of Governmental Experts regarding the legal, ethical, and humanitarian implications of autonomous weapon systems,

Acknowledging the potential benefits of military AI in intelligence analysis, logistics, surveillance, and decision-support systems,

1. Strengthening the implementation and clarification of existing International Humanitarian Law (IHL) obligations in the context of AI-enabled warfare, especially on protecting civilians and minimizing incidental civilian harm, through measures including:
    a. Establishing standardized international testing and evaluation procedures for AI-enabled military systems, through:
        i. Encouraging each Member State to designate a competent national testing authority to assess the reliability and limitations of military AI systems before deployment,
        ii. Encouraging national testing authorities to evaluate the potential for system errors, misidentification, and unintended civilian harm,
        iii. Recommending that national authorities conduct additional evaluations when significant modifications are made to an AI system,
        iv. Encouraging regular reviews of deployed systems when substantial changes are made to their software, capabilities, operational environment, or intended use;
    b. Supporting compliance with international humanitarian law, including distinction, proportionality, and precautions in attack, through:
        i. Establishing procedures requiring the modification, restriction, or suspension of deployment when legal reviews identify risks that prevent reliable compliance with applicable international law, and to conduct renewed legal reviews when significant changes to an AI system's software, capabilities, intended use, or operating conditions,
        ii. Inviting the International Committee of the Red Cross (ICRC) to provide technical and legal expertise on the humanitarian implications of military AI, within its mandates,
        iii. Promoting national legal-review authorities to evaluate whether the design, intended use, and foreseeable operating conditions of AI-enabled weapons allow compliance with IHL, including the principles of distinction and precautions in attack, while assessing risks of misidentification, unreliable AI outputs, unexpected system behavior, and potential civilian harm;

2. Encourages Member States to maintain meaningful human control, in accordance with International Humanitarian Law (IHL), over the development, deployment, and operation of AI-enabled military systems to minimize the risk of unintended attacks, civilian casualties, and unintended escalation, but not limited to:
    a. Ensuring that qualified human officials retain the authority and practical ability to intervene in, override, suspend, or deactivate AI-enabled military systems when necessary:
        i. Requiring accessible emergency override mechanisms for systems operating with significant autonomous capabilities,
        ii. Establishing clear-standardized procedures outlining when human intervention must occur during high-risk military operations,
        iii. Requiring each Member State to designate a responsible military commander or authorized human operator within its national chain of command to approve the use of lethal force,
    b. Requiring meaningful human review before the authorization of actions involving the use of lethal force, particularly when recommendations or targeting information have been generated by artificial intelligence, including:

i. Ensuring that human operators are provided with sufficient relevant information to independently assess AI-generated recommendations,
ii. Discouraging the blind acceptance of AI-generated recommendations without reasonable human verification,
iii. Ensuring Member States to provide adequate training for military personnel responsible for operating or supervising AI-enabled systems to identify potential AI errors, misidentification, unreliable outputs, and system failures;

c. Strengthening the power of the organization, such as the Convention on Certain Conventional Weapons (CCW) and the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE on LAWS), to improve international oversight of military AI, through:
i. Expanding the GGE on LAWS' role in developing internationally agreed standards for human control and accountability,
ii. Encouraging Member States to negotiate further measures addressing the risks of autonomous weapons systems,
iii. Promoting the consistent application of international humanitarian law to AI-enabled military systems;

3. Calls upon Member States to establish clear accountability mechanisms under United Nations to govern the development, testing, deployment, and operation of AI-enabled military systems, in order to prevent misuse, address violations of international law, and clarify the responsibilities of relevant parties, through measures such as, but not limited to:
a. Establishing clear chains of responsibility for individuals and institutions involved in the development and military use of AI-enabled systems, including:
i. Identifying the responsibilities of military commanders and operators responsible for authorizing or supervising the system's use,
ii. Establishing appropriate responsibilities for developers, manufacturers, and other relevant parties according to their level of involvement and applicable national and international law;
b. Requiring Member States to maintain and submit records regarding significant AI-enabled military operations, AI policies, safety procedures, and system failures, including:
i. The types of AI-based weapon systems developed or deployed,
ii. The human supervision procedures applied during their operation,
iii. The national legal review procedures conducted before deployment,
iv. Preserving relevant information necessary for subsequent investigations into suspected violations or system failures;
c. Recommending regular technical testing of AI-based weapon systems before deployment and after significant system modifications, including:
i. Evaluating the system's ability to distinguish military targets from civilians,
ii. Assessing system reliability under different operational conditions,
iii. Recording test results and addressing identified failures before authorizing deployment;
d. Encouraging Member States to exchange information on AI-related military incidents and safety failures through designated national reporting authorities, including, but not limited to:

i. Encouraging impartial and appropriate investigations when credible evidence suggests that an AI-enabled military system contributed to an unlawful attack or serious violation of international law,
ii. Exchanging information on identified technical failures and corrective measures,
iii. Submitting annual summaries of national implementation efforts to the relevant UN body;

4. Urges member states to develop the AISCBB (Artificial Intelligence Safety Cooperation Beyond Borders) facilitated by the United Nations Office for Disarmament Affairs (UNODA) to prioritize the spread of open-source AI safety tools that serve primarily to ensure the safe deployment of this technology, transparency, ethical development of AI's application without compromising sensitive data through ways such as, but not limited to:
    a. Establishing a voluntary international platform through which Member States may exchange non-sensitive information regarding military AI safety practices, including:
        i. Methods for identifying vulnerabilities, unreliable system behaviour, and potential risks of misidentification,
        ii. General safety standards and testing procedures that can be shared without revealing classified military capabilities,
        iii. Ensuring participation and engagement within the cooperation remains completely voluntary and that classified military information is not disclosed;
    b. Encouraging the collaborative development and distribution of open-source AI safety and evaluation tools, provided that such tools are designed primarily for defensive, testing, verification, or safety purposes, including:
        i. Tools capable of identifying unreliable outputs, data bias, or unexpected system behaviour,
        ii. Testing frameworks capable of evaluating whether AI systems perform reliably under changing or unpredictable conditions,
        iii. Holding regular international meetings, facilitated by the UNODA, to allow Member States to share updates on their military AI policies and safety,
    c. Establishing voluntary expert working groups involving representatives from Member States, AI researchers, and relevant international organizations to identify common challenges, develop shared terminology, and propose practical recommendations for improving transparency between countries and responsible AI governance,
    d. Encourages the development of standardized reporting guidelines for Member States to enhance the consistency and usefulness of information exchanged regarding military AIs:
        i. Developing a standardized-reporting template covering national AI policies, human oversight procedures, testing standards, and accountability mechanisms by UNODA, in consultation with Member States and relevant technical experts,
        ii. Establishing a regulation that highly encourages Member States to follow the reporting template to provide periodic updates on national approaches to the development and deployment of military AI, while excluding classified and sensitive information,
        iii. Inviting Member States to share non-sensitive information on identified risks, lessons learned, and improvements to their military AI safeguards,

e. Spreading awareness on the potential risks and issues AI-enabled autonomy has on contemporary warfare and public security, and distinguishing the differences between different aspects of AI's impact on contemporary warfare, distinguishing between AI-autonomous, autonomous, and AI-assisted – thus reducing ignorance amongst the public regarding this issue through a blend of conventional and contemporary methods including, but not limited to;
    i. Utilizing online and offline conventional methods such as news programs, posters, and physical newspapers to spread this message to rural communities and members of the demographic who lack digital and technological literacy,
    ii. Utilizing online contemporary methods to spread the urgency and importance of AI through social media, the use of 'trending' influencers to assist in spreading information, and integration into the modern education system to ensure the general public is well-educated at an early age;

5. Regulating the unethical development of autonomous AI Military Decision-Support Systems and Automated Defensive Systems to reduce the risks of an unstable arms race, reduce civilian casualties, and ensure officials maintain control over such systems to prevent potential decisions made due to AI misidentification in ways such as, but not limited to:
    a. Encouraging comprehensive safety testing and evaluation before systems with significant autonomous capabilities are deployed in operational environments, including:
        i. Testing systems against inaccurate, incomplete, biased, or deliberately manipulated information,
        ii. Evaluating how systems respond when sensors, communications, or relevant data become unreliable or unavailable,
        iii. Requiring developers to provide technical documentation on system limitations and known risks to the relevant military authorities,
    b. Encouraging Member States to establish clear minimum reliability and safety requirements for high-risk autonomous military systems before deployment, including:
        i. Requiring systems to demonstrate predictable behaviour within their intended operational environments,
        ii. Requiring appropriate fail-safe mechanisms capable of limiting or suspending operations when serious system failures are detected,
    c. Discouraging the deployment of systems in situations where their operators cannot reasonably understand their capabilities, limitations, or potential risks, including:
        i. Requiring relevant personnel to receive sufficient information regarding known limitations and potential failure scenarios,
        ii. Requiring additional testing or review when significant changes are made to the system's autonomous or AI-enabled capabilities,
    d. Developing international and mutually agreed-upon standards for the implementation of artificial intelligence in contemporary warfare through ways such as, but not limited to:
        i. Defining and developing criterions to define and classify use as either "AI-assisted", "Autonomous", or "AI-autonomous" to reduce ambiguity regarding the technology,
        ii. Establishing procedures for investigating incidents involving unintended civilian harm caused by AI-based weapon systems,

iii. Requiring relevant authorities to document investigation findings and identify failures in system operation or human supervision;

6. Encourages developed countries with greater technological capabilities to support developing countries in accessing and developing AI safety knowledge, policies, implementation, and technical expertise to ensure that all Member State can have equal access by:
    a. Promoting the support from developed countries by allowing them to get accessed to shared information,
    b. Ensuring participation of Member States to be completely voluntary and only share non-sensitive information that must not lead to security concerns.