

Position of the Russian Federation on the Prospects of Addressing the Issue of Military Use of Artificial Intelligence in the Military Domain at the International Level, Including Within the United Nations Framework

The Russian Federation attaches great importance to the issue of military application of artificial intelligence (AI) and its consideration at the international level. Our principled approaches to this matter are set out in the national contribution to the UN Secretary-General's report 80/78 on the opportunities and challenges in the area of maintaining international peace and security in connection with the use of AI in the military domain.

Parameters of dialogue on "military" AI

In general, we do not object to establishing a multilateral dialogue on the use of AI in the military domain and studying the implications of utilizing relevant means for international peace and security as proposed in the aforementioned report of the UN Secretary-General (para 45).

In this regard, we assume that such dialogue must be built on the basis of consensus and inclusivity and be of intergovernmental nature. The involvement of non-governmental actors should be exercised without prejudice to this process, be advisory and depoliticized in nature and remain strictly within the competence of the respective stakeholders.

These fundamental principles are aimed at ensuring that the interests of every Member State are taken into account, as well as at securing effectiveness of the decisions adopted. The relevance of such approach is further reaffirmed in the report of the UN Secretary-General, which, inter alia, underscores the need to conduct work on the basis of consensus (para 54), participation of all States in the deliberations on "military" AI (paras 48, 63, 66), as well as the exclusive prerogative of States in decision-making (para 48).

Platform for dialogue on AI in the military domain

We regard the UN disarmament mechanism as an effective and inclusive platform. It should play a central role in future discussions on the application of AI in the military domain within the World Organization's framework, as noted in the UN Secretary-General's report (para 52). Owing to its universal membership, deliberative mandate and specialized nature of work, the UN Disarmament Commission (UNDC) is ideally suited for such dialogue. The UNDC's extensive experience in developing recommendations on various aspects of arms control, disarmament and non-proliferation (ACDN), combined with the Commission's developments in the field of the use of AI in the military domain, would ensure constructive discussions and agreement on practical guiding principles and recommendations on the use of AI for military purposes.

It is of paramount importance that the issues of "military" AI considered within the UN, do not duplicate the topics discussed in relevant forums – most notably within the Group of Governmental Experts (GGE) of the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW) on emerging technologies in the area of lethal autonomous weapons systems (LAWS). Consideration within the GGE of issues related to the application of AI in the military domain has a comprehensive scope, is not confined strictly to LAWS and deals with a number of important aspects (including those of legal, technical and military nature) related to the use of these technologies for military purposes.

Given the GGE's track record of interaction, its ability to maintain a reasonable balance between humanitarian concerns and the legitimate defence interests of States as well as the participation of key States developing these technologies, the Group is regarded as the optimal platform for substantive work to align positions on the mentioned aspects of "military" AI (paras 47, 50 of the UN Secretary-General's report). It is important to continue this work within the GGE under a new mandate after 2026.

We consider it impractical to explore the options mentioned in the report of the UN Secretary-General (paras 52, 54) regarding the possible consideration of

“military” AI issues at the Conference on Disarmament (CD) or within a specialized mechanism, including a new relevant Open-Ended Working Group. Any attempt to “enrich” the agenda of the CD with a new item would only further aggravate the situation at the forum that has long been hampered by a lack of progress on its program of work. The initiation of a specialized process would require negotiating its organizational modalities and allocating additional financial resources from the UN budget, which appears impractical given liquidity crisis the World Organization is currently facing.

At the same time, we note that formats with limited membership promoting unilateral approaches to “military” AI, including summits on “responsible application” of these technologies (REAIM, mentioned in paras 41, 42 of the UN Secretary-General’s report), are of destructive nature. They are intended to substitute a genuine intergovernmental discussion in specialized forums and to elaborate, within a narrow circle, certain understandings and standards in this field, which would subsequently be imposed on the international community as ready-made solutions for adoption. Regional initiatives on the use of AI in the military domain (para 43 of the UN Secretary-General’s report) involve a limited number of participants, and their work objectively cannot be universal in nature. The outcomes of meetings in such formats with a limited composition do not reflect the positions of all interested States and cannot be considered as a basis for future work. Furthermore, such initiatives could result in fragmentation of efforts and undermine inclusive work on the issue within the designated relevant forums.

Looking ahead, we believe it is essential that discussions on “military” AI compliment existing dialogue formats on ACDN, and consider it necessary to avoid creating arbitrary international oversight or discriminatory export control mechanisms (para 32 of the UN Secretary-General’s report).

International Legal Regulation of AI Application in the Military Domain

At present the international community is at the very beginning of discussions within relevant multilateral forums on the application of new technologies in the military domain. The terminology, including the definition of

“AI” itself, requires further harmonization. Until a consensual “basic language” is elaborated, starting discussions on restrictions and prohibitions of the respective weapons systems would be premature. The opinion on the need to avoid premature adoption of regulatory measures is recorded in para 46 of the UN Secretary-General’s report. We also believe that the introduction of restrictions impeding legitimate innovation and technological progress (para 47 of the UN Secretary-General’s report) is unacceptable.

Analysis of the existing international legal framework related to the military aspects of AI application is a promising topic for the dialogue on this issue. Currently, the overwhelming majority of Member States agree that the norms of international law in general, and international humanitarian law (IHL) in particular, are applicable to weapons systems using AI, and that there is no strict prohibition on the use of such means in international law (para 28 of the UN Secretary-General’s report). Existing international law, including IHL, contains a broad range of restrictions that ensure humanization of the conduct of hostilities.

States should proceed from the fact that their military application of AI must comply with the norms of international law, as well as a shared understanding of how the international law applies to the use of AI technologies in the military domain should be established. A unified and clear understanding of how exactly existing international legal norms and principles are implemented in practice with regard to the “military” AI is a necessary precondition for the States to fulfill their obligations.

Meanwhile, it is critical to understand that AI technologies are of neutral nature and that the dialogue should rather focus on aspects related to the application of AI in the military domain (para 27 of the UN Secretary-General’s report) than on the technologies themselves. Furthermore, relevant international legal framework must not hinder technological development or restrict equal access to AI for developing countries.

The concept of “responsible application” of the “military” AI, outlined in paras 33-35 of the UN Secretary-General’s report, cannot be viewed as a viable

basis for comprehensive global discussion on the mentioned topic. This concept relies on principles that lack international consensus. Standards for “responsible application” do not constitute a universally recognized category of international law. They are subjective in nature and fail to clarify who determines, and by what criteria, whether a State’s use of these technologies is “responsible” or “irresponsible”. The mentioned concept serves to consolidate the leading positions of a certain group of countries in this field by imposing on the international community rules and standards that are profitable only for those particular States, which would give them a complete freedom of action while constraining the development of related technologies in “undesirable” countries. This position is shared by other States as well (para 46 of the UN Secretary-General’s report).

Integrating ethical considerations and international human rights law (IHRL, paras 28 and 57 of the UN Secretary-General’s report) into the dialogue is also counterproductive. These aspects cannot serve as an absolute and sufficient basis for establishing any kind of restrictive or prohibitive regimes on the mentioned types of weapons. Introducing them into the debate would only complicate the discussion related to the practical analysis of IHL, given existing disputes regarding its interaction with IHRL.

We consider it inappropriate to introduce into the deliberations such concepts as “reliability, explainability and traceability to prevent unintended consequences and malfunctions” (para 36 of the UN Secretary-General’s report). Such categories are subjective, non-legal in nature, and only complicate States’ efforts to forge a common understanding regarding the weapons systems and military equipment with the use of AI.

Advantages of weapons systems and military equipment with the use of AI

The report of the UN Secretary General outlines the advantages of technologies with the use of “military” AI, inter alia, those contributing to better compliance with international law, including IHL.

Speed is recognized as a main advantage of AI, including in information analysis and in decision-making. Other opportunities include “reliability, safety

and robustness” (para 7 of the UN Secretary General’s report), as well as increasing efficiency, accuracy and precision of the operations (para 56).

AI allows to conduct efficient analysis of large data sets, facilitate the detection of threats, enable increased situational awareness and more accurate operations as well as reduce risks to civilians and strengthen protection to civilian facilities (para 8). AI also reduces risks for military personnel on the frontlines as military systems using AI technologies can replace human when conducting military operations in remote or dangerous areas (para 12). AI optimizes administration and logistics as well as the predictive maintenance of complex systems (para 11 of the report).

AI is also capable of improving coordination and communication (para 9 of the UN Secretary-General’s report), enhancing information and communications technology security by detecting intrusions or other malicious activities, including for protection of critical infrastructure. Moreover, such technologies can facilitate the process of identifying content used for misinformation and disinformation (para 10 of the report).

AI contributes to the observance of norms of IHL, in particular its fundamental principles of distinction, proportionality and precautions in attack, as well as to the protection of civilians and civilian facilities. AI is able to improve situational awareness in general and understanding of the civilian environment in particular, as well as to increase accuracy and reduce the risk of human error (para 13 of the report).

Risks of “military” AI and their mitigation

In general, we do not object to analyzing advantages associated with using AI in the military domain along with the risks related to the application of such technologies within the framework of the relevant international dialogue, as proposed in para 46 of the UN Secretary-General’s report.

One of the objectives of international law is to minimize, as much as possible, negative consequences of an armed conflict by limiting the scope of means and methods of conducting combat operations available to its parties. These

requirements, however, apply to humans rather than to machines. Therefore, one of the elements in minimizing potential risks associated with the use of AI technologies in the military domain is the compliance in good faith by States with international law, including IHL.

Among the key provisions of IHL aimed at reducing the risks associated with new types of weapons (including “military” AI technologies), Article 36 of Additional Protocol I to the Geneva Conventions of 1949 is of particular significance. It stipulates that in the study, development, acquisition or adoption of new types of weapons, means and methods of warfare, a State must determine whether their employment would be subject to the prohibitions and restrictions provided for in relevant international treaties. The aforementioned Article does not contain provisions clarifying how such reviews should be conducted and does not impose an obligation on States to provide information on the results. At the same time the Article does not exclude the possibility of a voluntary exchange of best practices in implementing those reviews. This allows for a flexible and effective approach which takes into account national specifics and the positions of States, their capabilities and limitations in the technological field.

Maintaining human control related to the use of AI technologies in the military domain is aimed at addressing potential challenges associated with the military use of AI technologies. Its specific forms are laid down by States at the national level and are not necessarily ensured exclusively through direct control. Among possible methods of maintaining human control are increasing reliability and fault tolerance of the system; limiting types of targets; duration of operation, geographical coverage and scale of application; timely intervention and deactivation; testing in real operational environment; admitting persons who have successfully learned respective procedures to operation of relevant systems (control); controlling the production process of separate elements and the product as a whole; controlling disassembly and disposal of separate elements and the product as a whole.

A comprehensive exchange of views on relevant platforms, first and foremost within the GGE on LAWS, is also intended to minimize the challenges associated with “military” AI, inter alia through adoption of specific recommendations and guiding principles. During the sessions of this Group, the States, using specific examples, repeatedly emphasized the role of AI technologies in the military domain in fulfilling defense missions, combating terrorist threat and mine danger, etc.

The mentioned means of maintaining human control can effectively reduce risks, including those mentioned in the UN Secretary General’s report (paras 16, 18, 19). The methods of reducing the risks associated with “military” AI technologies presented in the report (para 37) are also relevant.

Topics for further discussion

In addition to the proposals already outlined in the present working paper on the application of the current international legal framework to “military” AI, as well as the risks and advantages of the respective technology, a number of issues mentioned in para 49 of the UN Secretary General’s report could also be considered as possible topics for further dialogue. In particular, it is essential to hold discussions on ensuring observance of international law, including IHL; developing a common understanding of definitions and terminology; transparency and confidence-building measures; the issue of autonomy in the use of force; AI systems that directly support combat operations; ensuring adequate data management mechanisms; strengthening international cooperation and assistance; supporting capacity-building in order to overcome the digital and the AI gaps, including through exchange of knowledge, transfer of technology and best practices. The issue of a clear distinction between “civil” and “military” AI should also be considered, as proposed in para 61 of the UN Secretary-General’s report.

In a possible exchange of views on international cooperation, special attention should be attached to ensuring a non-discriminatory and equal access to AI technologies for all States, especially countries of the Global South. Access to technologies for development purposes should not be restricted. Possible exchange

of information and experience in the context of the proposed discussion on transparency measures regarding the use of “military” AI should be carried out by States exclusively on a voluntary basis.

The proposal to ensure at the international level “assistance to national regulation, including in order to ensure private sector compliance with international law” (para 49 of the UN Secretary-General’s report) is counterproductive. The internal regulation of practical issues related to the observance of international law in the field of “military” AI, including by the private sector, should remain at the discretion of the States. It is inappropriate to discuss the aspects of “protecting human dignity and human rights” and “promoting continued regional dialogue”, mentioned in para 49 of the UN Secretary-General’s report, given the understandings presented in the previous paragraphs of the working paper.

It is crucial that the possible list of topics for dialogue to be formulated on the basis of the instruments supported by all States. For instance, a number of countries distanced themselves from or voted against the documents mentioned in the report of the UN Secretary-General such as the “Pact for the Future” (UNGA resolution 79/1, para 38), UNGA resolutions 78/241, 79/62 “LAWS” (para 39), as well as resolution 79/239 “AI in the military domain and its implications for international peace and security” (para 38). The promotion of those topics as the basis for international cooperation on AI will complicate constructive work.

The above understandings are intended to facilitate further effective and substantive discussion of the use of AI in the military domain at international level, including within the UN. We expect that the proposals of the Russian Federation will be fully taken into account, including when preparing summary of discussions on the outcome of the informal consultations pursuant to UNGA resolution 80/58.