

**Конвенция о запрещении или ограничении
применения конкретных видов обычного
оружия, которые могут считаться наносящими
чрезмерные повреждения или имеющими
неизбирательное действие**

7 March 2023

Russian only

**Группа правительственных экспертов
по новым технологиям в сфере создания
смертоносных автономных систем вооружений**

Женева, 6–10 марта и 15–19 мая 2023 года

Пункт 5 повестки дня

Активизация рассмотрения предложений и разработка на основе консенсуса возможных мер, в том числе принимая во внимание пример существующих протоколов в рамках Конвенции, и других вариантов, касающихся нормативной и оперативной основы по новым технологиям в сфере создания смертоносных автономных систем вооружений, основываясь на рекомендациях и выводах Группы правительственных экспертов по вопросам, касающимся новых технологий в сфере создания смертоносных автономных систем вооружений, и привлекая экспертные знания по правовым, военным и технологическим аспектам

**«О Концепции деятельности Вооруженных Сил
Российской Федерации в сфере разработки и применения
систем вооружений с использованием технологий
искусственного интеллекта»**

Рабочий документ Российской Федерации

Введение

26 июля 2022 года Министром обороны Российской Федерации утверждена Концепция деятельности Вооруженных Сил (ВС) Российской Федерации в сфере разработки и применения систем вооружений с использованием технологий искусственного интеллекта (ИИ).

Настоящий документ имеет целью закрепить принципы развития и применения систем вооружений с использованием технологий ИИ в ВС России, обеспечить соблюдение обязательств по международному праву, в том числе международному гуманитарному праву (МГП), а также заложить основы взаимодействия и сотрудничества с иностранными партнерами в сфере ИИ.

Положения настоящего документа надлежит использовать на всех этапах жизненного цикла систем вооружений с использованием технологий ИИ.

1. В документе развернуто представлен понятийный аппарат, применяемый как разработчиками, так и конечными пользователями систем вооружений с использованием технологий ИИ.
2. Концепция закрепляет систему взглядов и основные принципы в отношении систем вооружений с использованием технологий искусственного интеллекта:

Безопасность — недопустимость использования систем в целях умышленного причинения вреда гражданам и юридическим лицам, предупреждение и минимизация рисков возникновения негативных последствий использования технологий; при разработке или приобретении систем вооружений с использованием технологий ИИ



будут учитываться требования обеспечения физической защищенности, надлежащих нефизических гарантий (включая киберзащищенность от взлома или подмены данных), риск распространения, в том числе приобретения террористическими группами.

Прозрачность — объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей продуктов, которые созданы с использованием ИИ, к информации о применяемых в этих продуктах алгоритмах работы ИИ.

Технологический суверенитет — обеспечение необходимого уровня самостоятельности Российской Федерации в области искусственного интеллекта, в том числе посредством преимущественного использования отечественных технологий ИИ и технологических решений, разработанных на основе ИИ.

Ответственность — сохранение ответственности человека за решения по созданию и применению систем вооружений с использованием технологий искусственного интеллекта на всех этапах их жизненного цикла, а также в соответствии с применимыми нормами международного права, в том числе посредством обеспечения функционирования таких систем в рамках ответственной цепи командования и управления.

Контроль — необходимость сохранения контроля человека над системой вооружений с использованием технологий ИИ; конкретные формы и методы такого контроля остаются на усмотрение государства; эффективный контроль человека над системой может обеспечиваться не только путем прямого управления.

3. Особое внимание уделено вопросу соблюдения положений международного права, в том числе МГП, а также ответственности за их осуществление.

Подтверждено, что международное право, включая Устав Организации Объединенных Наций, МГП, полностью применимы ко всем оружейным системам, в частности к разработке и задействованию систем вооружений с использованием технологий ИИ.

Обязательным является соблюдение основных принципов МГП — различие, соразмерность, гуманность и военная необходимость, а также проведение правового обзора при изучении, разработке, приобретении или принятии на вооружение новых видов оружия. Не допускается применение систем вооружений с использованием технологий искусственного интеллекта запрещенными МГП способами (методами) ведения боевых действий. При этом гражданское население и комбатанты постоянно остаются под защитой международного права.

В Концепции закрепляется необходимость сохранения контроля человека над машиной для обеспечения выполнения действующих норм международного права. Такой контроль осуществляется за счет:

- повышения надежности и отказоустойчивости;
- ограничения типов целей;
- ограничения по продолжительности функционирования, географическому охвату и масштабу применения;
- своевременного вмешательства и деактивации;
- проведения испытаний систем вооружений с использованием технологий ИИ в реалистичных оперативных условиях;
- допуска к управлению (контролю) лиц, успешно освоивших процедуры ответственного применения систем вооружений с использованием технологий ИИ;
- контроля процесса производства отдельных элементов и изделия в целом;
- контроля разборки и утилизации отдельных элементов и изделия в целом.

Подтверждено, что ответственность за соблюдение обязательств по международному праву, а также применение систем вооружений с использованием технологий ИИ ложится на должностное лицо, которое ставит задачу таким системам вооружений и отдает приказ на их применение.

4. Согласно Концепции, взаимодействие с иностранными государствами в сфере разработки и применения систем вооружений с использованием технологий ИИ осуществляется в соответствии с требованиями международного права и национального законодательства, включая нормы режимов экспортного контроля, сохранности государственной тайны и интеллектуальной собственности, и с соблюдением обоюдной выгоды от реализации соответствующих совместных программ.

В документе подтверждено, что Конвенция о «негуманном» оружии представляет собой оптимальную площадку для рассмотрения проблематики смертоносных автономных систем вооружений с учетом ее уникального характера.

5. Концепция представляет потенциальные преимущества использования систем вооружений с использованием технологий ИИ. Одновременно дана оценка возможных рисков применения таких систем. К таким рискам относятся:

- попадание систем вооружений с использованием технологий ИИ в руки негосударственных субъектов, включая террористические формирования;
- потеря контроля над системой вследствие технического сбоя или ее взлома и перепрограммирования злоумышленниками;
- принятие системой или оператором ошибочных решений.

В целях минимизации вышеуказанных рисков в отношении систем вооружений с использованием технологий ИИ предусматриваются следующие меры:

- эффективное управление жизненным циклом;
- организация разносторонних испытаний на всех этапах жизненного цикла, в том числе в условиях, близких к реальной обстановке;
- обеспечение надежности и отказоустойчивости;
- определение критериев готовности;
- обеспечение максимальной защищенности от несанкционированного доступа;
- подготовка операторов;
- приоритет в использовании технологий ИИ в процессе сбора и обработки информации, необходимой для поддержки принятия решений в военной сфере;
- непрерывный контроль за действиями таких систем со стороны оператора, обеспечение возможности экстренного прекращения выполнения боевой задачи по его команде;
- исключение попадания негосударственным субъектам, которые могут использовать их в противоправных целях.

Оценка рисков и меры по их уменьшению должны быть частью цикла проектирования, разработки, испытания и развертывания новых технологий в любых оружейных системах.

Заключение

Ответственное применение вооружений, в том числе с использованием технологий ИИ, является одним из приоритетов Вооруженных Сил Российской Федерации.

Крайне важно на всех этапах минимизировать риски причинения непредвиденного ущерба. Важно развивать систему экспортного контроля с целью

предотвращения попадания систем вооружений с использованием технологий ИИ в руки негосударственных субъектов, которые могут их использовать в злонамеренных целях. Как следствие, необходимо сохранение контроля за такими системами на всех этапах их жизненного цикла.

Особое значение в этом контексте приобретает налаживание добровольного обмена опытом и передовыми достижениями между государствами в данной сфере.

Обязательным условием при разработке систем вооружений с использованием технологий ИИ является учет требований МГП.

Реализуя настоящие концептуальные взгляды, Вооруженные Силы Российской Федерации будут стремиться к использованию технологий искусственного интеллекта в интересах обороноспособности государства при строгом соблюдении норм международного права.

[Unofficial English version]

Concept of Activities of the Armed Forces of the Russian Federation in the Development and Use of Weapons Systems with Artificial Intelligence Technologies

Submitted by the Russian Federation

Introduction

The Minister of Defence of the Russian Federation approved the Concept of Activities of the Armed Forces of the Russian Federation in the development and use of weapons systems with artificial intelligence (AI) technologies on July 26, 2022.

That document aims to consolidate the principles of development and use of weapons systems with AI technologies in the Armed Forces of the Russian Federation, ensure compliance with the obligations under international law, including international humanitarian law (IHL), and lay the foundation for interaction and cooperation with foreign partners in the field of AI.

The provisions of that document should be used at all stages of the life cycle of weapons systems with AI technologies.

1. The document expands on the conceptual framework used by both developers and end users of weapons systems with AI technologies.
2. The concept enshrines a system of views and basic principles related to weapons systems with AI technologies:

Security – it should be inadmissible to use such systems for the purpose to intentionally harm individuals and entities, to aware and mitigate risks of negative consequences of the use of technologies; development or acquisition of weapons systems with AI technologies should take into account physical security requirements, appropriate non-physical safeguards (including cyber- security against hacking or data spoofing), risk of proliferation, including acquisition by terrorist groups.

Transparency – the work of artificial intelligence and process of achieving results by AI should be explainable; users of the products created with the help of artificial intelligence should enjoy non-discriminatory access to the information about AI algorithms used in these products.

Technological sovereignty – the necessary level of independence of the Russian Federation in the field of artificial intelligence should be ensured, including through the predominant use of domestic AI technologies and AI-based solutions.

Responsibility – human should be retained responsible for decisions to develop and use weapons systems with artificial intelligence technologies throughout their life cycle, as well as in accordance with applicable norms of international law, including by ensuring operation of such systems within a responsible command and control chain.

Control – need to maintain human control over a weapon system with AI technologies; specific forms and methods of such control should be left to the discretion of the state; effective human control over system can be achieved beyond direct control.

3. Compliance with the provisions of international law, including IHL, as well as responsibility for their implementation, receives special attention.

The Concept recognizes that international law, including the UN Charter and IHL, continues to fully apply to all weapons systems, including the development and use of weapons systems with AI technologies.

Adherence to the basic principles of IHL: distinction, proportionality, humanitarian considerations and military necessity should be mandatory, as well as legal review in the study, development, acquisition or adoption of new weapon types. Weapons systems with artificial intelligence technologies should be outlawed in IHL prohibited means or method of warfare. At the same time, civilians and combatants at all time remain under the protection of the international law.

The Concept enshrines the need to maintain human control over machine to uphold compliance with the existing norms of international law. Such control should be exercised by means of:

- increased reliability and fault tolerance;
- limitations on types of targets;
- limitations on duration of operation, geographic scope and scale of use;
- timely intervention and deactivation;
- testing of weapons systems with AI technologies in real operation environment;
- admitting persons who have successfully mastered the procedures of responsible use of weapons systems with AI technologies to management (control);
- control over the production process of separate elements and the product as a whole;
- control over presorting and disposal of separate elements and the product as a whole.

The document recognizes that responsibility for compliance with obligations under international law, as well as the use of weapons systems with AI technologies should lie with the official who assigns a task to such weapons systems and gives an order to use such weapons systems.

4. According to the Concept, cooperation with foreign countries in the development and use of weapons systems with AI technologies should be carried out in accordance with the requirements of international law and national legislation, including the norms of export control regimes, the protection of state secret and intellectual property, as well as with the respect for mutual benefit from the implementation of relevant joint programmes.

The document recognizes that the Convention on Certain Conventional Weapons offers an appropriate framework for dealing with the issue of emerging technologies in the area of lethal autonomous weapons systems given Convention's unique nature.

5. The Concept presents the potential benefits of weapons systems with AI technologies, while providing an assessment of possible risks posed by the use of such systems. Such risks include:

- falling of weapons systems with AI technologies into the hands of non-state actors, including terrorist entities;
- loss of control of the system due to a technical failure or hacking and reprogramming by perpetrators;
- making erroneous decisions by system or operator.

To mitigate the above mentioned risks, the following measures are envisaged for weapons systems with AI technologies:

- effective life cycle management;
- comprehensive tests at all life cycle stages, including in realistic environment;
- ensuring reliability and fault tolerance;
- specification of readiness criteria;
- ensuring maximum security against unauthorized access;
- operator training;

- prioritizing the use of AI technologies in gathering and processing information necessary to support military decision-making;
- continuous operator monitoring of the actions of such systems, ensuring emergency termination of a combat mission by operator's command;
- preventing access by non-state actors who could use them for illegal purposes.

Risk assessment and mitigation measures should be part of the new technologies design, development, testing and deployment cycle in any weapon system.

Conclusion

Responsible use of weapons, including those with AI technologies, is one of the priorities of the Armed Forces of the Russian Federation.

It is essential to minimize the risks of unintended damage at all stages. It is important to develop the export control system to prevent weapons systems with AI technologies from falling into the hands of non-state actors who might use them for malicious purposes. As a consequence, it is necessary to maintain control over such systems throughout their entire life cycle.

The voluntary exchange of relevant best practices and cutting-edge technological achievements between states is of particular importance in this context.

Respect for the requirements of IHL is a prerequisite for the development of weapons systems with AI technologies.

By implementing these conceptual views, the Armed Forces of the Russian Federation will seek to use artificial intelligence technologies in the interest of national defence capability, in strict compliance with the norms of international law.
