

Robotics and AI Regulatory Research Centre



www.robopravo.ru

Model Convention on Robotics and Artificial Intelligence

Rules for the Creation and Use of Robots and Artificial Intelligence

Version 1.0. November 2017

Authors: Andrey Neznamov & Viktor Naumov

PREAMBLE

We, the parties to this convention, whereas:

(1) in this day and age, humanity faces the global task of determining the paths for its development in the coming decades;

(2) one of the most pressing tasks of our time is to comprehend the role of progress in science and technology, ascertain the opportunities it offers and assess the risks and consequences of its impact upon the global order, the nation state, the economy, society and the individual;

(3) in recent decades, progress in science and technology has seen the intensive development and ubiquitous application of cyber-physical systems with a variety of uses, including robots;

(4) humanity is placing great hope in the development of robotics and artificial intelligence, which have a demonstrable capacity to address many of the problems that we have accrued and to give a new impetus to the development of the global community;

(5) as this new landscape comes into view, the potential dangers and threats that might arise by reason of the use of the new technologies of robotics and artificial intelligence cannot be ignored;

(6) we cannot rule out scenarios in which the use, en masse, of robots and artificial intelligence could have catastrophic consequences for the existing global order and the entire human species;

(7) despite the growth in the number of robots and cyber-physical systems, the global community as a whole has hitherto not developed universal concepts regarding the rules for the interaction that their use gives rise to;

(8) it cannot but be a source of concern that, to date, no understandings have been reached or rules put in place regarding the use of high-risk robots that, by virtue of their design and purpose, are capable of causing harm to a great number of people;

(9) many people's ideas of human interaction with the intelligent robot are created primarily by popular culture and are often limited to Isaac Asimov's laws of robotics, which can lead to the misperception that these laws will be ubiquitously incorporated into the design of all cyber-physical systems;

(10) although some nation states have begun in recent years to develop legislation governing robots, international law and the majority of national legal systems do not provide special models and mechanisms for regulating societal relations as concern the use of robots; and

(11) as matters now stand, there is a clear necessity for an effective consolidation of efforts by all interested nation states, innovative businesses, the global scientific community and experts to work out common approaches and legal and ethical standards in the sphere of the creation, development, incorporation, use and proliferation of cyber-physical systems,

have come to the understanding that we must work together to create this Convention — an 'open-source' document; all interested parties are invited to be involved in the discussion and improvement of it.

The goal of this Model Convention is to identify the main problems that might arise in society and the legal system by reason of the intensive development of cyber-physical systems; to bring together the array of fundamental rules of robotics that have been expressed at various times by various experts; to plot possible directions for finding solutions to existing and/or anticipated problems; and to initiate discussion between experts in different countries in order to develop a unified set of ethical and legal concepts applicable in rules for the creation and use of robots and artificial intelligence.

INTRODUCTORY PROVISIONS

Article 1. Agents in robotics

The participants in the processes of creating, incorporating or using robots and cyber-physical systems with artificial intelligence are the researchers, developers and funders of research and development in the sphere of robotics and those who produce, own, possess or operate robots, governmental authorities and any regulatory authorities and users and other human beings that have interaction with robots and cyber-physical systems, inter alia systems with artificial intelligence.

Article 2. Objects in robotics

In robotics, objects include all categories of robots, in the broadest possible sense, whatever their purpose and level of risk, mobility and automation, and cyber-physical systems with artificial intelligence in any form. Unless otherwise expressly arises from the text of this Model Convention, the term 'robot' should hereinafter be construed in the broadest possible sense and, for the purposes of this Model Convention, as also including robotic mechanisms and cyber-physical systems, inter alia those with artificial intelligence.

Article 3. Terminology

'Robot' and derivatives of it, such as 'robotic mechanism' and 'intelligent robot', are to be defined according to the universally accepted meanings of these terms in the country in question and/or in relation to the specific model of robot. If these terms are not defined, existing international standards can be used, in particular ISO 8373:2012 'Robots and robotic devices – Vocabulary'.

Article 4. Universality of rules

The rules for the creation, incorporation and use of robots are addressed to all agents in robotics.

Article 5. National provisions

Participants in the processes of creating, incorporating and using robots have an obligation to know and comply with statutory requirements in force in the place in which robots are in use or are planned to be used ("applicable law") and to take account of other applicable rules, including ethical and religious norms.

RULES FOR ROBOT SAFETY

Article 6. Principle of robot safety

The creation, incorporation and use of robots must be carried out in a way that ensures, to the maximum extent, safety for the individual, society and the state, and without any harm being caused to the individual, humanity, living creatures and all of their habitat.

Article 7. Prohibition upon the initiation of harm by robots

Unless this Model Convention establishes otherwise, robots may not be created that are capable of, at their own initiative, deliberately causing harm to a human being; criteria, processes and consequences relating to the causing of damage to a person's property will be determined by applicable law.

Article 8. Restriction of harm by intelligent robots

Intelligent robots must not cause harm to a human being or humanity or, through inaction, allow such harm to come about, unless the causing of harm is necessary and justified in order to avert greater harm. 'Harm' will be defined based on the information that the intelligent robot has available to it.

Article 9. Rule of equal risk

In any processes carried out using robots, a human being must not be exposed to a risk of harm to his or her life or health greater than the risk that he or she would have faced in the process in question had it not involved robots.

Article 10. Awareness of robot safety

All information regarding the danger that a robot poses to a human being, society or the environment must be in the public domain, free of charge and easily accessible to any agent in robotics.

Article 11. Protection for confidential information

Access to information gathered and processed by robots and the use of it must not infringe personal privacy and/or the rules for safeguarding other types of confidential information under applicable law.

Article 12. Protection for robots against unauthorised access

Robots must be fitted with a system that protects them from unauthorised physical or electronic access to their systems and mechanisms.

Article 13. High-risk robots

Stricter requirements for protection from unauthorised access by third parties must be established for the operation of robots potentially capable of causing a human being substantial, inter alia fatal harm by virtue of their design and purpose (medical robots, implantable devices, highly-automated vehicles, military robots etc.). Agents in robotics must proceed on the basis of the assumption that such robots are a hazard and the competent authorities must establish the requirements in question and oversee compliance with them on the territory within their jurisdiction.

Article 14. Operation of high-risk robots

Any action by high-risk robots that potentially has negative consequences for a human being because of their design and purpose may be carried out only through direct operation by another human being.

Article 15. Controllability of robots by a human being

Robots must, as far as practicable and justifiable, according to the particularities of the purpose of the type of robot in question, be directly or indirectly controllable by a human being.

Article 16. 'Black box'

Robots must at all times record and store data about their functioning and all action being taken by them ('black box'). Access to that data must be given to the human beings responsible for the robots' actions and proper functioning and to the competent authorities, through the processes established by applicable law.

Article 17. 'Red button'

Robots that physically interact with human beings and are not under the direct control of a human being must have an instantaneous or emergency shut-down function that operates on demand ('red button').

GENERAL RULES FOR THE CREATION OF ROBOTS

Article 18. Creation of the common good

Developers and producers of robots and other human beings involved in the process of creating robots ('creators of robots') must strive as far as possible to ensure that the robots created by them maximise the good for as many human beings as possible.

Article 19. Responsible attitude towards consequences of incorporation of robots

The creators of robots and funders of research and development in the sphere of robotics must understand that their activities are not purely technical in nature, take a responsible attitude to the social and economic phenomena and circumstances that might arise as a result of the incorporation of robots and strive to avert any serious negative consequences associated with them.

Article 20. Priority of safety

In the creation of robots, the objective of ensuring the safety of the human being must at all times prevail over all other objectives and tasks.

Article 21. Feasibility of compliance with Model Convention

The creators of robots must procure that it is functionally feasible for the rules in this Model Convention (as corrected, amended and supplemented) to be complied with.

GENERAL RULES FOR THE USE OF ROBOTS

Article 22. Observance of human rights and universally accepted standards of morality and ethics

Fundamental human rights and universally accepted standards of morality and ethics must be observed in the process of using robots, irrespective of whether there is specific legal regulation in place.

Article 23. Respect for human dignity

A human being must not debase human dignity by his or her treatment of robots.

Article 24. Prevention of unauthorised harm by a robot

A human being with the ability to prevent or stop the unauthorised (inter alia accidental) harming of another human being by a robot must take all possible steps to do so that do not involve a risk to his or her own life or health.

Article 25. Identification of robots

The information that a device or object is a robot must be communicated to any human being or other robot interacting with it, except where otherwise dictated by circumstance or where it is not necessary based on the purpose and specific conditions of use of the form of robot in question.

Article 26. Restriction on robot behaviour

Robots must function in a way that is fundamentally predictable to their creators and users, corresponds to their design and purpose and is safe and controllable.

Article 27. Awareness of robot functioning

Robots must function in such a way that agents in robotics interacting with them can understand the processes by which the robots are functioning or can obtain sufficient information about this at the time of the interaction.

**Article 28. Responsibility for compliance
with rules for robot functioning**

Unless otherwise flows from the provisions of applicable law or the specific circumstances, responsibility for compliance with the rules for robot functioning will lie with the creators of robots and any other human being who can affect compliance with them by his or her actions.

Article 29. Robots as agents in law

Robots may act in civil commerce as independent persons, inter alia may act as the owners of other robots, if applicable law expressly so provides.

RULES FOR THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

Article 30. Responsible attitude

Developers, researchers and funders of research in the sphere of artificial intelligence and related fields must take account of public sentiment and are not entitled to disregard the view that artificial intelligence technologies might have consequences that are irreversible for humanity and carry with them existential risks.

Article 31. Presumption of danger of artificial intelligence

Developers, researchers and funders of research in the sphere of artificial intelligence and related fields must proceed on the basis of a presumption that artificial intelligence is dangerous, meaning that an artificially intelligent technology that has been or is going to be created is dangerous to human beings until proven otherwise.

Article 32. Conscious interaction

There must be no interaction between a human being and artificial intelligence without the human being's awareness and consent.

Article 33. Application of the provisions of the Convention to artificial intelligence

The provisions in this Model Convention concerning robots should be applied to cyber-physical systems with artificial intelligence, taking into account the particularities of their design.

RESTRICTIONS UPON THE USE OF MILITARY ROBOTS

Article 34. Applicability of humanitarian rules for the conduct of war

Robots may not be used for military purposes in breach of the humanitarian rules for the conduct of war that are universally accepted by the international community.

Article 35. Restrictions for the purposes of complying with humanitarian rules for the conduct of war

Robots being created for military purposes must have unmodifiable restrictions built into them at the outset that correspond to the provisions of international humanitarian law restricting the methods and means of conducting war, regardless of where the robots are to be deployed.

Article 36. No harm to civilian population

Robots being deployed for military purposes must not be used to cause harm to the civilian population. When robots are used for military purposes, all steps possible under the circumstances must be taken to preclude or minimise harm by robots to the life or health of the opponent's personnel.

Article 37. Responsibility for crimes committed using robots

Responsibility for war crimes committed by robots will be determined according to the rules that apply to war crimes committed by human beings controlling the robots. Nation states must strive to put in place provisions whereby the use of robots in the commission of war crimes will be treated as an aggravating factor.

DEVELOPMENT OF THE RULES FOR ROBOTICS AND ARTIFICIAL INTELLIGENCE

Article 38. Voluntary application

Until such time as the provisions of this Model Convention become binding, agents in robotics may state that they will apply all or some of the provisions of this Convention voluntarily.

Article 39. Scope of use of Model Convention

The rules in this Convention may be used in any jurisdiction and by any community of participants in the processes of creating, incorporating and using robots, as a foundation for the unification of rules for the creation and use of robots.

Article 40. Cooperation in the development of universally accepted international rules

Agents in robotics must facilitate international cooperation in the development of universally accepted rules for the creation, incorporation and use of robots and the harmonisation of standards for robotics such that their application is practicable in as many jurisdictions as possible.

Article 41. Assistance in creation of transnational institution

Intergovernmental and non-governmental institutions must be created, inter alia under the auspices of existing international associations and organisations, to enable the effective, harmonious and safe development of robotics and artificial intelligence.

Article 42. Development of Model Convention

The provisions of this Convention may be discussed, corrected, amended and supplemented in order that, to the maximum extent possible, they reflect current conceptions of robots and artificial intelligence in human society and are in keeping with universally accepted rules and standards, which must have effect in the sphere of robotics and artificial intelligence.