Guiding principles for General-Purpose Robots

Modern technological advancements enable the creation of highly intelligent robots that closely resemble humans in appearance and behavioral characteristics (hereafter referred to as general-purpose robots or GPRs). These robots potentially capable of performing a wide range of tasks historically carried out by humans, from hazardous industrial work and customer service to applications in medical services.

Interest in the robotics industry, particularly in general-purpose robots, is growing significantly every year, raising new questions and concerns. General-purpose robots and humanoid robots have the potential to radically transform various sectors of the economy, but their widespread adoption brings up important ethical issues that require special attention. Among them there are issues of privacy and security, the impact on employment and workforce displacement, the level of robot autonomy, and the distribution of responsibility for their actions.

These and other aspects shall be taken into account when constructing and operating general-purpose robots to ensure their ethical use and positive impact on society and individuals.

These Guiding principles for General-Purpose Robots (hereafter referred to as the Guiding principles) are developed in addition to and based on the AI Ethics Code created by the AI Alliance Russia. The Guiding principles provide recommendations and guidance for all stakeholders on adopting an ethical approach to the design, creation, and application of general-purpose robots (GPRs) in alignment with moral and ethical principles.

I. General Provisions

1. Scope and Application of the Guiding Principles

1.1. The Guiding principles serve as an industry-specific document designed to specify the provisions of the Code of Ethics for AI as applied to the field of GPRs and, in particular, humanoid robots.

1.2. The Guiding principles provide general ethical principles and recommendations aimed at ensuring ethical practices throughout all stages of the life cycle of general-purpose robots and related technologies.

1.3. The recommendations outlined in these Guiding principles apply exclusively to general-purpose robots intended for civilian (non-military) purposes.

1.4. Adherence to the Guiding principles by those who develop, implement, use, and supply general-purpose robots (hereafter referred to as GPR Actors) is voluntary.

1.5. The provisions of the Guiding principles may be expanded and/or specified for certain groups of GPR Actors through the development of separate addendums and annexes. These documents will account for technological advancements, the nature of tasks performed, the class and purpose of general-purpose robots, the level of associated risks, as well as the specific context and environment in which GPRs are applied.

1.6. The provisions of these Guiding principles are advisory in nature.

2. Objectives of the Guiding Principles

The primary objective of the Guiding principles is to establish conditions for the:

2.1. protection of human rights and freedoms at all stages of the life cycle of general-purpose robots;

2.2. ensurance of safe and ethical development, production, deployment, use, and disposal of general-purpose robots;

2.3. prevention of negative consequences of the use of general-purpose robots;

2.4. advancement science and technology in the field of robotics and general-purpose robots;

2.5. support of innovation and the competitiveness of the domestic robotics industry, including fostering Russian developers of general-purpose robots;

2.6. strengthening public trust in the use of general-purpose robots;

2.7. increase of public awareness and understanding of the capabilities and limitations of robots;

2.8. encouragement of collaboration between developers, manufacturers, users, and owners of robots;

2.9. establishment of ethical standards and principles to guide all participants in the robotics market;

2.10. ensurance that the scenarios and methods of the GPR utilization comply with the legislation of the Russian Federation and international norms.

3. Terms and Definitions

The terms and definitions are based on national standard of the Russian Federation GOST R 60.0.0.4-2019/ISO 8373:2012 «Robots and robotic devices. Terms and definitions».

3.1. Robot – an executive mechanism that is programmable in two or more degrees of mobility, possesses a certain degree of autonomy, and is capable of moving within its external environment to perform intended tasks.

3.2. General-Purpose Robot (GPR) – a multifunctional universal robot designed to perform a wide range of tasks in specified conditions. GPRs have autonomy, mobility, environmental perception, and the ability to interact with humans.

3.3. Autonomy – the ability to perform tasks as intended based on the robot's current state and perception of its external environment without human intervention.

3.4. Anthropomorphic (humanoid) Robot – a robot with a body, head and limbs that resemble human form and movement.

3.5. GPR Actors:

3.5.1. Developers – individuals or organizations responsible for designing and developing general-purpose robots.

3.5.2. Manufacturers – companies that produce components or systems for general-purpose robots.

3.5.3. Users – operators, consumers, and organizations that use general-purpose robots to perform specific tasks.

3.5.3.1. Operator – an individual who initiates, monitors, and terminates the operations performed by the robot.

3.5.3.2. Consumer – a person interacting with a robot to gain some form of benefit.

3.5.4. Owners – individuals or organizations that own general-purpose robots.

II. Ethical Principles and Recommendations in the Field of General-Purpose Robots

4. Human-Centered Approach

4.1. The development, implementation, and utilization of general-purpose robots shall prioritize the rights and legitimate interests of individuals and society as a whole.

4.2. Interaction with robots should not adversely affect a person's ability for self-fulfillment, work, or intellectual activity. Robots shall not assume functions and roles that individuals wish to retain for themselves. The decision as regards to which aspects of life to entrust to a robot must always remain with the individual.

4.3. Equal opportunities and access to the use of GPRs must be ensured for all individuals, regardless of race, gender, age, nationality, religion, or social status.

4.4. Efforts shall be taken to ensure fairness and equality in the distribution of benefits associated with the use of GPR.

4.5. GPR actors should apply ethical principles at all stages of the lifecycle of general-purpose robots.

4.6. GPR shall be designed in a way to minimize the emergence of negative emotions (e.g., fear due to the «uncanny valley» effect) during human-robot interactions.

4.7. Developers are encouraged to incorporate algorithms into GPRs that align with the cultural characteristics and spiritual and moral values of the peoples of the Russian Federation. GPR should be designed to display ethical behavior when interacting with humans.

4.8. The development and use of GPR for satisfying sexual needs and/or needs for emotional intimacy, as well as for creating the illusion of genuine intimacy between humans and GPR, are prohibited.

4.9. Designers and manufacturers of GPRs should take the measures necessary to avoid excessive human dependence on the robot.

5. Non-Maleficence Principle

5.1. Ensuring the safety and well-being of humans during interactions with GPR is a fundamental principle in their development, deployment, and use.

5.2. The use of GPR for purposes that violate the legislation of the Russian Federation is strictly prohibited. This includes violations of individual rights and legitimate interests, causing physical, psychological, or emotional harm to people or other living beings, and damaging property, except in cases specified in clause 7.2.

5.3. Developers shall employ algorithms and implement measures that prevent GPR from performing unlawful and/or unethical actions that could result in harm to humans, other living beings, or property.

5.4. Developers and manufacturers shall provide for safety measures and protocols that prevent GPR from executing unlawful or unethical instructions from users, particularly those capable of causing harm to humans, other living beings, or property.

6. General-Purpose Robots and the Right to Employment

6.1. When implementing GPRs, it is necessary to assess the impact of their use on the right to employment as provided for in the legislation. An evaluation of the risks of violating people's right to work shall be conducted in each specific case.

6.2. The implementation of robots shall be carried out in a balanced manner, taking into account the rights and interests of employees.

6.3. Employers are encouraged to take reasonable and sufficient measures to provide timely retraining for employees who may be replaced by general-purpose robots. As a best practice, employers are advised to support such employees in finding further employment and/or provide other forms of support, including financial assistance.

6.4 Employers are encouraged to refrain from abusing automation of work processes that could violate employees' right to work and/or result in discriminatory practices by the employer.

6.5. When introducing GPRs, employers are encouraged to adopt approaches and foster a culture of collaborative interaction between employees and robots.

7. Robots as a Problem-Solving Tool

7.1. General-purpose robots serve as tools for solving routine tasks for humans, with the scope of such tasks potentially defined by industry-specific documents. In this context, GPRs are considered property and assets of their owner or, in the case of a Robot-as-a-Service (RaaS) model, the property of the service provider company.

7.2. GPR operators must handle robots with due care during their use. Intentional harm to robots is unacceptable, as it may lead to risks of projecting such behavior onto humans and could infringe upon the rights of third parties, including ownership rights related to GPRs. Exceptions may apply in situations where actions that could harm a robot are necessary to prevent and/or stop harm to a human, another living being, or property.

8. Accountability

8.1. General-purpose robot technologies must always remain under human control, and only humans should have the necessary authority to make final decisions in cases where such decisions involve substantial risks to human rights, legitimate interests, or public interests.

8.2. General-purpose robots may possess a certain degree of autonomy. However, developers should design GPRs in such a way that they always operate strictly within the framework of predefined and lawful instructions (tasks) provided by the owner and/or operator of the GPR.

9. Safety

9.1. GPR actors are advised to implement comprehensive information security measures in compliance with current security standards. These measures should include internal mechanisms to protect general-purpose robots from unauthorized interference, including through malicious software.

Users of GPRs should be informed about the risks of such interference and the rules of information security when using general-purpose robots.

9.2. Developers and manufacturers of GPRs are recommended to establish a system and protocols for early warning of security risks, as well as a rapid response plan for addressing such breaches.

9.3. Developers of GPRs should ensure the ability to perform an emergency shutdown of the robot, including its complete deactivation, in cases of loss of effective control over the robot, threats to human life and/or health, risks of property damage, or other socially dangerous consequences. Measures should be in place to minimize potential negative outcomes of an emergency shutdown, such as preventing a robot from falling abruptly, which could harm people or property, and addressing other possible negative effects.

9.4. The shutdown function for GPRs can be implemented as an automatic deactivation, manual shutdown, or remote-control deactivation. In any case, such a mechanism shall be accessible to the user of the GPR.

10. Transparency

10.1. Transparency and interpretability of artificial intelligence algorithms used in general-purpose robots (GPRs) must be ensured to the extent necessary to guarantee that the actions of GPRs are understandable to users. Relevant information should be communicated to users within the limits under the confidentiality and protected information regime.

10.2. Comprehensive information about the capabilities and limitations of GPRs shall be made available to users, enabling them to make informed and conscious decisions about the use of general-purpose robots.

10.3. GPRs should be programmed to provide full information about their functions, capabilities, and limitations upon user request. They should also be capable of explaining their actions, intentions, and decision-making logic in a language understandable to the user.

10.4. It is recommended to include error-handling subsystems in GPRs to notify users of malfunctions. These notifications should be supplemented with recommendations for troubleshooting.

10.5. Users shall be provided with the ability to provide feedback by creating specific channels for submitting complaints, inquiries, and suggestions to improve the performance of GPRs.

10.6. To uphold the principle of transparency during the development and production of GPRs, comprehensive documentation shall be provided. This documentation should describe in detail the AI algorithms used in such robots, the data they collect and utilize, and other relevant information regarding their creation, production, and utilization. Developers should also provide descriptions of GPR operation scenarios in various circumstances.

11. Privacy and Personal Data Protection

11.1. GPR actors shall program robots to ensure the confidentiality of the collected and processed data.

11.2. The development and programming of robots must adhere to the principle of privacy-by-design (data protection and confidentiality by default) and strictly comply with the legislation of the Russian Federation on personal data protection.

11.3. Users must be informed about what data the GPR processes, where it is stored, and what security measures are applied to protect it. User data must be protected by default.

12. Non-Discrimination

12.1. The actions of GPRs must be free from biases, prejudices, or stereotypes of developers, as well as from practices or behaviors that discriminate against any groups of people.

12.2. Under no circumstances should GPRs contribute to the proliferation of discriminatory practices. On the contrary, developers and manufacturers should take cultural, religious, and other differences into account when designing, developing, and producing GPRs to ensure guarantees of ethical behavior by general-purpose robots.

13. Liability

13.1. Liability for harm caused by the actions of a general-purpose robot is assigned to GPR actors in accordance with the general principles of legal responsibility prescribed by the Russian Federation legislation.

13.2. Liability for the actions of a GPR may be imposed on any GPR actor, depending on the specific circumstances of the case.

13.3. The liability of GPR actors must correspond to the nature, degree, and extent of the harm that may result from the utilization of general-purpose robots. When determining the degree of fault, the following factors should be considered: the role of the GPR actor in the robot's lifecycle, their influence on the robot’s actions, the potential and actual impact of the actor on the harm caused, the extent of the damage, and the actor's real ability to prevent such harm or damage.

14. Collaboration

14.1. GPR actors shall prioritize social responsibility and sustainability in the development and application of general-purpose robot technologies, contributing to the comprehensive advancement of this field.

14.2. To foster the development of the general-purpose robot industry, promote technological progress, minimize risks, and prevent adverse outcomes, GPR actors are encouraged to:

14.2.1. promote international collaboration with GPR developers and manufacturers to exchange knowledge, experience, and technologies in robotics and general-purpose robot development;

14.2.2. strengthen engagement with the international scientific community and international organizations to advance and improve GPR technologies and coordinate efforts in the field of GPR ethics;

14.2.3. collaborate with regulators and government authorities of the Russian Federation to develop a comprehensive policy framework for regulating the GPR industry;

14.2.4. cooperate with industry organizations to develop and implement harmonized standards and recommendations in robotics, as well as to create a rating system for reliable and responsible GPR manufacturers and developers;

14.2.5. deepen cooperation with public organizations and associations to consider the interests of various social groups in the development and use of GPRs;

14.2.6. facilitate effective interaction between manufacturers and users of GPRs to ensure high-quality maintenance and support for robots;

14.2.7. strengthen collaboration with developers from related industries and technologies to create a comprehensive technological ecosystem that fosters innovation and supports the sustainable development of the GPR industry.

15. Adherence to and Implementation of the Guiding principles

15.1. The mechanism for joining these Guiding principles and its implementation is carried out in accordance with the provisions outlined in subsection 2 of Section 2 of the AI Ethics Code.