Artificial Intelligence Code of Ethics

The Code of Ethics in the Field of Artificial Intelligence (hereinafter referred to as the Code) establishes the general ethical principles and standards of conduct that should be followed by participants in relation to the field of artificial intelligence (hereinafter referred to as AI Actors) in their activities, as well as the mechanisms for the implementation of the provisions of this Code.

The Code applies to relationships related to the ethical aspects of the creation (design, construction, piloting), implementation and use of AI technologies at all stages that are currently not regulated by the legislation of the Russian Federation and/or by acts of technical regulation.

The recommendations of this Code are designed for artificial intelligence systems (hereinafter referred to as AIS) used exclusively for civil (not military) purposes.

The provisions of the Code can be expanded and/or specified for individual groups of AI Actors in industry-specific or local documents on ethics in the field of AI, considering the development of technologies, the specifics of the tasks being solved, the class and purpose of the AIS and the level of possible risks, as well as the specific context and environment in which the AIS are being used.

SECTION I

PRINCIPLES OF ETHICS AND RULES OF CONDUCT

1. THE MAIN PRIORITY OF THE DEVELOPMENT OF AI TECHNOLOGIES IS PROTECTING THE INTERESTS AND RIGHTS OF HUMAN BEINGS COLLECTIVELY AND AS INDIVIDUALS

1.1. Human-centered and humanistic approach.

In the development of AI technologies, the rights and freedoms of the individual should be given the greatest value. AI technologies developed by AI Actors should promote or not hinder the realization of humans’ capabilities to achieve harmony in social, economic and spiritual spheres, as well as in the highest self-fulfillment of human beings. They should take into account key values such as the preservation and development of human cognitive abilities and creative potential; the preservation of moral, spiritual and cultural values; the promotion of cultural and linguistic diversity and identity; and the preservation of traditions and the foundations of nations, peoples and ethnic and social groups.

A human-centered and humanistic approach is the basic ethical principle and central criterion for assessing the ethical behavior of AI Actors, which are listed in the section 2 of this Code.

1.2. Respect for human autonomy and freedom of will.

AI Actors should take all necessary measures to preserve the autonomy and free will of a human‘s decision-making ability, the right to choose, and, in general, the intellectual abilities of a human as an intrinsic value and a system-forming factor of modern civilization. AI Actors should, during AIS creation, assess the possible negative consequences for the development of human cognitive abilities and prevent the development of AIS that purposefully cause such consequences.

1.3. Compliance with the law.

AI Actors must know and comply with the provisions of the legislation of the Russian Federation in all areas of their activities and at all stages of the creation, development and use of AI technologies, including in matters of the legal responsibility of AI Actors.

1.4. Non-discrimination.

To ensure fairness and non-discrimination, AI Actors should take measures to verify that the algorithms, datasets and processing methods for machine learning that are used to group and/or classify data concerning individuals or groups do not intentionally discriminate. AI Actors are encouraged to create and apply methods and software solutions that identify and prevent discrimination based on race, nationality, gender, political views, religious beliefs, age, social and economic status, or information about private life. (At the same time, cannot be considered as discrimination rules, which are explicitly declared by an AI Actor for functioning or the application of AIS for the different groups of users, with such factors taken into account for segmentation)

1.5. Assessment of risks and humanitarian impact.

AI Actors are encouraged to assess the potential risks of using an AIS, including the social consequences for individuals, society and the state, as well as the humanitarian impact of the AIS on human rights and freedoms at different stages, including during the formation and use of datasets. AI Actors should also carry out long-term monitoring of the manifestations of such risks and take into account the complexity of the behavior of AIS during risk assessment, including the relationship and the interdependence of processes in the AIS’s life cycle.

For critical applications of the AIS, in special cases, it is encouraged that a risk assessment be conducted through the involvement of a neutral third party or authorized official body when to do so would not harm the performance and information security of the AIS and would ensure the protection of the intellectual property and trade secrets of the developer.

2. NEED FOR CONSCIOUS RESPONSIBILITY WHEN CREATING AND USING AI

2.1. Risk-based approach.

The level of attention to ethical issues in AI and the nature of the relevant actions of AI Actors should be proportional to the assessment of the level of risk posed by specific technologies and AISs and the interests of individuals and society. Risk-level assessment must take into account both the known and possible risks; in this case, the level of probability of threats should be taken into account as well as their possible scale in the short and long term. In the field of AI development, making decisions that are significant to society and the state should be accompanied by scientifically verified and interdisciplinary forecasting of socio-economic consequences and risks, as well as by the examination of possible changes in the value and cultural paradigm of the development of society, while taking into account national priorities.

In pursuance of this Code, the development and use of an AIS risk assessment methodology is recommended.

2.2. Responsible attitude.

AI Actors should have a responsible approach to the aspects of AIS that influence society and citizens at every stage of the AIS life cycle. These include privacy; the ethical, safe and responsible use of personal data; the nature, degree and amount of damage that may follow as a result of the use of the technology and AIS; and the selection and use of companion hardware and software.

In this case, the responsibility of the AI Actors must correspond to the nature, degree and amount of damage that may occur as a result of the use of technologies and AIS, while taking into account the role of the AI Actor in the life-cycle of AIS, as well as the degree of possible and real impact of a particular AI Actor on causing damage, as well as its size.

2.3. Precautions.

When the activities of AI Actors can lead to morally unacceptable consequences for individuals and society, the occurrence of which the corresponding AI Actor can reasonably assume, measures should be taken to prevent or limit the occurrence of such consequences. To assess the moral acceptability of consequences and the possible measures to prevent them, Actors can use the provisions of this Code, including the mechanisms specified in Section 2.

2.4. No harm.

AI Actors should not allow use of AI technologies for the purpose of causing harm to human life, the environment and/or the health or property of citizens and legal entities. Any application of an AIS capable of purposefully causing harm to the environment, human life or health or the property of citizens and legal entities during any stage, including design, development, testing, implementation or operation, is unacceptable.

2.5. Identification of AI in communication with a human.

AI Actors are encouraged to ensure that users are informed of their interactions with the AIS when it affects their rights and critical areas of their lives and to ensure that such interactions can be terminated at the request of the user.

2.6. Data security

AI Actors must comply with the legislation of the Russian Federation in the field of personal data and secrets protected by law when using an AIS. Furthermore, they must ensure the protection and protection of personal data processed by an AIS or AI Actors in order to develop and improve the AIS by developing and implementing innovative methods of controlling unauthorized access by third parties to personal data and using high-quality and representative datasets from reliable sources and obtained without breaking the law.

2.7. Information security.

AI Actors should provide the maximum possible protection against unauthorized interference in the work of the AI by third parties by introducing adequate information security technologies, including the use of internal mechanisms for protecting the AIS from unauthorized interventions and informing users and developers about such interventions. They must also inform users about the rules regarding information security when using the AIS.

2.8. Voluntary certification and Code compliance.

AI Actors can implement voluntary certification for the compliance of the developed AI technologies with the standards established by the legislation of the Russian Federation and this Code. AI Actors can create voluntary certification and AIS labeling systems that indicate that these systems have passed voluntary certification procedures and confirm quality standards.

2.9. Control of the recursive self-improvement of AISs.

AI Actors are encouraged to collaborate in the identification and verification of methods and forms of creating universal ("strong") AIS and the prevention of the possible threats that AIS carry. The use of "strong" AI technologies should be under the control of the state.

3. HUMANS ARE ALWAYS RESPONSIBILITY FOR THE CONSEQUENCES OF THE APPLICATION OF AN AIS

3.1. Supervision.

AI Actors should provide comprehensive human supervision of any AIS to the extent and manner depending on the purpose of the AIS, including, for example, recording significant human decisions at all stages of the AIS life cycle or making provisions for the registration of the work of the AIS. They should also ensure the transparency of AIS use, including the possibility of cancellation by a person and (or) the prevention of making socially and legally significant decisions and actions by the AIS at any stage in its life cycle, where reasonably applicable.

3.2. Responsibility.

AI Actors should not allow the transfer of rights of responsible moral choice to the AIS or delegate responsibility for the consequences of the AIS's decision-making. A person (an individual or legal entity recognized as the subject of responsibility in accordance with the legislation in force of the Russian Federation) must always be responsible for the consequences of the work of the AI Actors are encouraged to take all measures to determine the responsibilities of specific participants in the life cycle of the AIS, taking into account each participant's role and the specifics of each stage.

4. AI TECHNOLOGIES SHOULD BE APPLIED AND IMPLEMENTED WHERE IT WILL BENEFIT PEOPLE

4.1. Application of AIS in accordance with its intended purpose.

AI Actors must use AIS in accordance with the stated purpose, in the prescribed subject area and for solving the prescribed problems.

4.2. Stimulating the development of AI.

AI Actors should encourage and incentivize the design, implementation, and development of safe and ethical AI technologies, taking into account national priorities.

5. INTERESTS OF DEVELOPING AI TECHNOLOGIES ABOVE THE INTERESTS OF COMPETITION

5.1. Correctness of AIS comparisons.

To maintain the fair competition and effective cooperation of developers, AI Actors should use the most reliable and comparable information about the capabilities of AISs in relation to a task and ensure the uniformity of the measurement methodologies.

5.2. Development of competencies.

AI Actors are encouraged to follow practices adopted by the professional community, to maintain the proper level of professional competence necessary for safe and effective work with AIS and to promote the improvement of the professional competence of workers in the field of AI, including within the framework of programs and educational disciplines on AI ethics.

5.3. Collaboration of developers.

AI Actors are encouraged to develop cooperation within the AI Actor community, particularly between developers, including by informing each other of the identification of critical vulnerabilities in order to prevent their wide distribution. They should also make efforts to improve the quality and availability of resources in the field of AIS development, including by increasing the availability of data (including labeled data), ensuring the compatibility of the developed AIS where applicable and creating conditions for the formation of a national school for the development of AI technologies that includes publicly available national repositories of libraries and network models, available national development tools, open national frameworks, etc.

They are also encouraged to share information on the best practices in the development of AI technologies and organize and hold conferences, hackathons and public competitions, as well as high-school and student Olympiads.

They should increase the availability of knowledge and encourage the use of open-knowledge databases, creating conditions for attracting investments in the development of AI technologies from Russian private investors, business angels, venture funds and private equity funds while stimulating scientific and educational activities in the field of AI by participating in the projects and activities of leading Russian research centers and educational organizations.

6. IMPORTANCE OF MAXIMUM TRANSPARENCY AND TRUTHFULNESS IN INFORMATION ON THE LEVEL OF DEVELOPMENT, CAPABILITIES

AND RISKS OF AI TECHNOLOGIES

6.1. Credibility of information about AIS.

AI Actors are encouraged to provide AIS users with credible information about the AIS, acceptable and most effective methods of using the AIS and the harm, benefits, and existing limitations of their use.

6.2. Raising awareness of the ethics of AI application

AI Actors are encouraged to carry out activities aimed at increasing the level of trust and awareness of citizens who use AISs and society in general. This should include increasing awareness of the technologies being developed, the features of the ethical use of AISs and other provisions accompanying the development of AIS. This promotion could include the development of journal articles, the organization of scientific and public conferences and seminars, and the inclusion of rules of ethical behavior for users and operators in the rules of the AIS.

SECTION 2

APPLICATION OF THE CODE

1. Foundation of the code action

1.1. Legal basis of the Code.

The Code takes into account the legislation of the Russian Federation,

the Constitution of the Russian Federation and other regulatory legal acts and strategic planning documents. These include the National Strategy for the Development of Artificial Intelligence, the National Security Strategy of the Russian Federation and the Concept for the Regulation of Artificial Intelligence and Robotics. The Code also considers international treaties and agreements ratified by the Russian Federation applicable to issues ensuring the rights and freedoms of citizens in the context of the use of information technologies.

1.2. Terminology.

Terms and definitions in this Code are defined in accordance with applicable regulatory legal acts, strategic planning documents and technical regulation in the field of AI.

1.3. AI Actors.

For the purposes of this Code, AI Actors is defined as persons, including foreign ones, participating in the life cycle of an AIS during its implementation in the territory of the Russian Federation or in relation to persons who are in the territory of the Russian Federation, including those involved in the provision of goods and services. Such persons include, but are not limited to, the following: developers who create, train, or test AI models/systems and develop or implement such models/systems, software and/or hardware systems and take responsibility for their design;

customers (individuals or organizations) receiving a product;

or a service; data providers and persons involved in the formation of datasets for their use in AISs; experts who measure and/or evaluate the parameters of the developed models/systems; manufacturers engaged in the production of AIS; AIS operators who legally own the relevant systems, use them for their intended purpose and directly implement the solution to the problems that arise from using AIS;

operators (individuals or organizations) carrying out the work of the AIS; persons with a regulatory impact in the field of AI, including the developers of regulatory and technical documents, manuals, various regulations, requirements, and standards in the field of AI; and other persons whose actions can affect the results of the actions of an AIS or persons who make decisions on the use of AIS.

2. MECHANISM OF ACCESSION AND IMPLEMENTATION OF THE CODE

2.1 Voluntary Accession.

Joining the Code is voluntary. By joining the Code, AI Actors agree to follow its recommendations.

Joining and following the provisions of this Code may be taken into account when providing support measures or in interactions with an AI Actor or between AI Actors.

2.2 Ethics officers and/or ethics commissions.

To ensure the implementation of the provisions of this Code and the current legal norms when creating, applying and using an AIS, AI Actors appoint officers on AI ethics who are responsible for the implementation of the Code and who act as contacts for AI Actors on ethical issues involving AI. These officers can create collegial industry bodies in the form of internal ethics commissions in the field of AI to consider the most relevant or controversial issues in the field of AI ethics. AI Actors are encouraged to identify an AI ethics officer whenever possible upon accession to this Code or within two months from the date of accession to the Code.

2.3. Commission for the Implementation of the National Code

in AI Ethics.

In order to implement the Code, a commission for the implementation of the Code in the field of AI ethics (hereinafter referred to as the Commission) being established. The commission may have working bodies and groups consisting of representatives of the business community, science, government agencies and other stakeholders. The Commission considers the applications of AI Actors wishing to join the Code and follow its provisions; it also maintains a register of Code members.

The activities of the Commission and the conduct of its secretariat are carried out by the Alliance for Artificial Intelligence association with the participation of other interested organizations.

2.4. Register of Code participants.

To accede to this Code, the AI Actor sends a corresponding application to the Commission. The register of AI Actors who have joined the Code is maintained on a public website/portal.

2.5. Development of methods and guidelines.

For the implementation of the Code, it is recommended to develop methods, guidelines, checklists and other methodological materials to ensure the most effective observance of the provisions of the Code by the AI Actors

2.6. Code of Practice.

For the timely exchange of best practices, the useful and safe application of AIS built on the basic principles of this Code, increasing the transparency of developers' activities, and maintaining healthy competition in the AIS market, AI Actors may create a set of best and/or worst practices for solving emerging ethical issues in the AI life cycle, selected according to the criteria established by the professional community. Public access to this code of practice should be provided.