Res Intell Manuf Assem, 2025, 4(1): 185-191
DOI: 10.25082/RIMA.2025.01.005

COMMENTARY

International Standardization Safe to Use of Artificial Intelligence

Evgeniy Bryndin

Research Centre "Natural Informatic", National Supercomputer Technological Platform, Novosibirsk, Russia

Correspondence to: Evgeniy Bryndin, Research Centre "Natural Informatic", National Supercomputer Technological Platform, Novosibirsk, Russia;
Email: bryndin15@yandex.ru

Received: March 6, 2025;
Accepted: May 4, 2025;
Published: May 9, 2025.

Citation: Bryndin E. International Standardization Safe to Use of Artificial Intelligence. *Res Intell Manuf Assem*, 2025, **4**(1): 185-191.
https://doi.org/10.25082/RIMA.2025.01.005

Abstract: Nowadays, the symbiosis of human abilities and the mastery of artificial intelligence will contribute to increased productivity and excellence in industry and social services. The use of artificial intelligence in various fields requires standardization of the safety of its knowledge and skills. International collaboration on artificial intelligence safety standardization is expanding. The UN has created a Global Advisory Body on Artificial Intelligence to support the efforts of the international community of specialists in managing intelligent systems related to the risks and safety of their use. The author proposes international standard of safe application of ensemble intelligent interoperable agents. Ensembles of agents with artificial intelligence are multi-agent synergistic self-organizing systems that function according to the laws of development, synergy and self-organization. Ensembles of intellectual agents solve the problem in the course of self-organization and cooperation according to the criteria of preference and restriction. The solution is considered found when, in the course of their nondeterministic interactions, agents reach the best consensus (temporary equilibrium or balance of interests), which is taken as a solution to the problem. The advantages of intelligent agents that allow you to build self-organizing ensembles are especially manifested in conditions of a priori uncertainty and high dynamics of the world around you, allowing you to build adaptive ensembles with communicative abilities, rebuilding your plans for events in real time. The higher the intelligence of each agent and the richer the opportunities for communication between agents, the more complex and creative behavior the ensemble can demonstrate. The intellect of the ensemble arises and manifests itself in the process of self-organization of intellectual agents. Intelligent agents use a physical, informal and logical model of the environment. That is, they use both attributes and sets of entities, processes, relationships, etc. Modern technologies allow you to create ensembles of intelligent agents with communication abilities, characterized by high openness, flexibility and efficiency, performance, scalability, reliability and survivability, approaching the intellectual abilities of a person and professional teams in their cognitive and functional capabilities and even sometimes surpassing them.

Keywords: artificial intelligence, security of intelligent systems, international cooperation

1 Introduction

The slightest errors in the design of intelligent systems can lead to catastrophic consequences. In Arizona, an unmanned car from Uber hit a woman crossing the street in the wrong place. In the driver's chair was a pilot, but he did not have time to stop the car. This accident was the first fatal accident involving a car with a third level of autonomy. It turned out that the laser radars of the car recognized the pedestrian as much as 5.6 seconds before the accident. But the algorithm decided not to reduce speed and began emergency braking only 0.2 seconds before the collision. All such facts in order for humanity, as far as possible, to act ahead of schedule, predict the possible dangers that may arise when introducing technologies using artificial intelligence. Experts say this today. Hazards are technological, legal, legal and ethical. New technologies pose both technical and ethical challenges. experts express various approaches to the principles of establishing responsibility for the actions of artificial intelligence: the responsibility of a particular subject - a manufacturer, developer, owner, user, expert or programmer. Human ingenuity and the desire for perfection, combined with the capabilities of new technologies, can solve the problems of mankind. Security solutions can ensure standardization of artificial intelligence. ISO focuses on standards relevant to the information and communication technology (ICT) industry. International Standards Association is focused on reaching out to government and industry in all of the locations around the world

where its governance meetings are held. There are currently no standards to guide ethical AI development and deployment, or to help consumers develop trust in AI. ISO pioneered work on standards related to ethically aligned design, this area is still in its infancy. The integration of AI enabled technologies in the daily lives of ordinary people is rapidly increasing. An appropriate standard could provide consumers with a reasonable level of comfort and assurance that AI has been developed conforming to ethical principals that protect their rights, e.g. privacy, transparency, and inclusiveness.

Standards Russia has recently formed a committee to study ethical AI and how they can map into existing international work on AI at ISO. 99 percent of people don't know how standards make modern society work. Standardization professionals, as well as those that understand the profession and its impact, are only one percent of the population. General population 99 percent expect everything to work, often with little interest in the details. They only notice when it does not work, and then it's a manufacturer or a government that are held to task when this happens (not standards).

Standards are mostly voluntary, with the ones that governments adopt become regulatory. By driving greater informed choice for consumers, there is heightened competition between developers and companies to gain market share in new areas so everything just works. Standardization in these areas will ensure that. If there is truly one percent that are aware of the impact, then this is indication of the huge responsibility that standards professionals have to benefit humanity to ensure everything works. The importance of standards to the work and careers of ICT practitioners continues to motivate the content of new innovative standardization activities to spark creativity and enthusiasm to solve safety problems. The standardization of artificial intelligence safety will help to find boundaries in which artificial intelligence will benefit humanity, not harm.

2 On the standardization of information security

Information security concerns the safety of artificial intelligence. The classic information security triad CIA is the most recognized and common in the international professional community. It was recorded in national and international standards and entered the main educational and certification programs for information security, such as CISSP and CISM.

Information security is responsible for the confidentiality, integrity and availability of information. In the concept of information security, specialists call them the principles of information security. Confidentiality means that only one who has the right to do this has access to information. The integrity means that the information is in full and does not change without the consent of the owner. Accessibility means that one who has the right to access information can get it.

Artificial intelligence specialists for information security mainly use the CIA triad. All three components: confidentiality, integrity and accessibility synonymically considered as principles, security attributes, properties, fundamental aspects, information criteria, the most important characteristics or basic structural elements. Certification, crypto protection and cybersecurity are also taken into account in the standardization of information security.

3 On international standardization of safe artificial intelligence

The coming years will take to increase safety and standardize the development and application of viable strong artificial intelligence. International standardization of the production and use of intelligent systems ensuring their compatibility has intensified.

The safety of artificial intelligence systems refers to an interdisciplinary field of research related to the prevention of accidents, their misuse and various harmful consequences that they can lead to, including technical problems, risk monitoring systems and high reliability. The security of artificial intelligence systems is necessary for smart factories, health centers, cafes, services, vehicles, agriculture, defense industry, etc.

The development of standard and criteria for the creation of systems with artificial intelligence that will be safe for humanity remains one of the urgent tasks.

The safety of the behavior of a system with artificial intelligence depends on its spatial, temporal, objective, visual and sound sensitivity within the boundaries of its use in the environment. The practical use of artificial intelligence systems in various spheres of society requires the

introduction of safety standards.

Safety for artificial intelligence and ethical codes on the use of intellectual systems are developed in a wide format of directions by specialists of various companies by different countries at the international level.

(1) The standardization of safe artificial intelligence in DeepMind was carried out in 2018. The safety of artificial intelligence systems was based on specifications, reliability and guarantees [1]. Specifications - guarantee that the behavior of the artificial intelligence system corresponds to the true intentions of the operator / user. Reliability - guarantees that the artificial intelligence system will continue to work safely at interference. Guarantees - give confidence that we are able to understand and control artificial intelligence systems during work.

(2) The AI Watch study is aimed at developing artificial intelligence safety standards for systems with minimal, limited and high levels of risk.

(3) For the safety of different artificial intelligence systems in Europe, ISO/IEC standards are developed:

ISO/IEC TR 24028: Information technology and artificial intelligence. The standard gives determination of the reliability of artificial intelligence systems, including approaches to establishing trust in artificial intelligence systems due to transparency, explanability and handling; Technical risks and threats to artificial intelligence systems, methods for mitigating the consequences of risks and threats are determined; Approaches to the assessment of failure tolerance, reliability, accuracy and safety and confidentiality.

ISO/IEC WD 5338: Information technology, artificial intelligence processes of life cycle of artificial intelligence. The standard is aimed at providing processes that support, control and improve artificial intelligence systems.

ISO/IEC AWI TR 5469: Artificial intelligence functional safety and artificial intelligence systems. The standard contains a description of the properties, risk factors, methods and processes of application and control of artificial intelligence in security systems.

ISO/IEC AWI TR 24368: Information technologies, artificial intellectual approaches and social services. The standard determines the ethical and social standards of artificial intelligence.

ISO/IEC AWI TR 24372: Information technologies, artificial intelligence - computing approaches to artificial intelligence systems. The standard determines modern computing approaches to artificial intelligence systems, computing characteristics, algorithms and methods, use options according to the ISO/IEC TR 24030 standard.

ISO/IEC CD 24668: Information technology, artificial intelligence structure of process management for big data analysis. The standard describes the reference model of the big data analysis process.

ISO/IEC WD TS 4213: Information technologies, artificial intelligence, assessments of machine learning classification. The standard is aimed at determining the methodology for measuring the effectiveness of classification models, systems and machine learning algorithms.

ISO/IEC 23894: Information technology, artificial intelligence, risk management. The standard provides recommendations for risk management that organizations face during the development and application of artificial intelligence methods and systems. In addition, the standard describes the processes of effective implementation and integration of risk management of artificial intelligence, which can be used in any organization.

ISO/IEC CD 38507: Information technologies, artificial intelligence management, consequences of using artificial intelligence systems. The standard provides a guide for organizations that use or consider the possibility of using artificial intelligence systems.

ISO/IEC WD 42001: Information technologies, artificial intelligence, management system. The standard is aimed at the formation of requirements and the creation of a guide to implement, maintain and improve artificial intelligence management systems in the context of a particular organization.

IEEE P2863: Standard for organization of control systems of artificial intelligence. The standard contains management criteria as security, transparency, accountability, responsibility and minimization of bias, as well as the stages of the process for effective implementation, audit of effectiveness, training and compliance with the development or use of artificial intelligence systems in organizations.

IEEE P3333.1.3: Standard for a deep assessment of visual experience based on the human factor. The standard determines the metric of content analysis of content and evaluating the quality of visual content based on deep training. The standard includes a description of deep learning models, visual perception indicators, virtual and mixed reality, clinical analysis and psychophysical data. The standard also includes images databases.

(4) Since 2021, the Code of Ethics of Artificial Intelligence has been operating in Russia. The code establishes the general ethical principles and standards of behavior that should be guided by participants in relations in the field of artificial intelligence in their activities. Russian experts have developed standards that regulate the safety of artificial intelligence systems not only for people, but also for the environment. Standardization concerns the introduction of artificial intelligence in various fields of human activity, such as transport, medicine, education, construction and a number of others. On September 30, 2023, the Russian Association, the House of Indo-Russian Technological Cooperation (Chamber for Russian Technology Collaboration, Cirtc) and the Russian Technical Committee No. 164 of the Rosstandart of the Russian Federation signed two memorandum of cooperation intentions aimed at developing relations between Russia And India in IT oblast. One of the documents concerns the standardization of artificial intelligence, as well as the creation of a joint laboratory for certification of solutions in the field of artificial intelligence. Interaction in the standardization of artificial intelligence will apply to the participants of the BRICS+. What will help to develop and apply the standards common to the BRICS countries. The Minister of Information Technology of India Rajiv Chandrakar proposed to develop a global security standard for artificial intelligence so that intellectual systems do not harm a person and social, industrial and natural environment.

(5) In 2023, the United States, Great Britain and more than ten other countries announced the signing of an international agreement on how to protect artificial intelligence systems. The document involves the creation of AI platforms designed in such a way that they are safe from the very beginning of their development.

(6) In 2023, representatives of 28 individual countries, including the USA, EU, Canada, China, Singapore, Japan, South Korea, Israel, India and the United Arab Emirates signed an international declaration for the safe use of artificial intelligence.

(7) Case for the use of strong artificial intelligence, developed by I. Ts. Natural formatics [2–8], approved by the Japanese Technical Committee for Standardization of Artificial Intelligence, is an international standard: a.111 Application of Strong Artificial Intelligence - "ISO/IEC JTC 1/SC 42 /WG 4 No 254 TR 24030 Working DRAFT V10 " - ISO/IEC 24030: 2019 (E). The case for the use of strong artificial intelligence contains generalized options. The standardization of the use of strong artificial intelligence ends with the developer by a specification of generalized options for each targeted use. The standard case contributes to the use of strong artificial intelligence in promising modern areas, such as: multi modal generative artificial intelligence, cooperation of intellectual digital doubles and humans, ethical artificial intelligence, quantum artificial intelligence, legitimization of artificial intelligence, intellectual chabitization, semantic emotional dialogue, and so on.

(8) The British Institute of Standards in 2024 introduced the global guide to the safety of artificial intelligence, which helps to responsibly use intellectual systems and manage them in companies around the world. The BSI standard for security eliminates key risks, ensuring the conformity of innovation to advanced experience. The standard for the safety of artificial intelligence is recommended for use in the field of services and in industry.

(9) In 2024, experts in the field of education and artificial intelligence of various countries develop international ethical standards for the use of intellectual systems for training. The standards of Japan provide for the use of generative tools of artificial intelligence in schools, from elementary grades to high school. On February 14, 2024, the National Research Institute for the Study of Generative Artificial Intelligence began to function in Japan. Japan began testing artificial intelligence systems in primary, junior and high schools. Japanese private companies have created several systems with artificial intelligence for Japanese schools. The Konica Minolta system is able to analyze students' reaction to the material presented, can collect data on the level of concentration of students, and activity in the rise of the hands. The system from Techno Horizon is designed to analyze the emotional state of each of the students. The artificial system helps to identify which children are excited, which children are in a state of stress or bored and concentrated children. Intellectual systems monitor the performance and effectiveness of the education of schoolchildren, give recommendations to teachers in the learning process.

4 International standard – Safe application of ensemble intelligent interoperable agents

Standard application of ensemble of intelligent interoperable agents defines parameters, characteristics, methods, models of digital double , knowledge, skills, behavior, images and other entities of intelligent virtual agent interaction (Table 1–7). Intelligent virtual agent interaction uses categorical method of utility and preference [9]. Synergetic mechanisms of self-organization, such as multi-level reflection, semantic and behavioral ontology, of technological ensembles of intelligent agents are basic for standardization when using ensembles in various fields [10]. Communicative-associative intelligent ensemble of diversified agents with an intelligent interface in the form of interacting AI assistants can implement a digital intelligent clinic [11].

Table 1 General

Use standard name	Safe application of ensemble of intelligent interoperable agents			
Application domain	Hi-Tech Labor Market			
Deployment model	Human digital double			
Status	Results of research: Strong Artificial Distributed Intelligence			
Scope	Economic and technical sectors and social services			
Objective(s)	Find accurate and universal application of strong artificial distributed intelligence			
Narrative	Short description (not more than 150 words)	Ensemble is complex of intelligent interoperable agents interacting through smart interface, implementing either technological process, social services, multi-inter- trans-disciplinary research, or production cycle.		
	Complete description	Ensemble is complex of intelligent interoperable agents interacting through smart interface, implementing either technological process, social services, multi-inter- trans-disciplinary research, or production cycle. In the ensemble, the whole range of tasks by certain rules is distributed among all agents. Job allocation means assigning each agent a role whose complexity is determined by the agent's capabilities. To organize the task distribution process, the ensemble creates either a distributed problem solution system or decentralized artificial intelligence. In the first version, the process of decomposition of the global problem and the inverse process of composition of the found solutions takes place under the control center agent. At the same time, the creative ensemble is designed strictly from top to bottom, based on the roles defined for the agents and the results of dividing the global task into subtasks. In the case of decentralized artificial intelligence, task distribution occurs during agent interaction and is synergistic.		
Stakeholders	Highly technological producer			
Stakeholders' assets, values	Reputation			
System's threats and vulnerabilities	Legal and ethical aspects of interaction with society.			
Key performance indicators (KPIs)	ID	Name	Description	Reference to mentioned use case objectives
	1	AI management of professional cooperation process	The technology of processes control can itself predict execution of certain stages on the basis of accumulated information about their labour intensity, selection of the route of agents and competences. Optimize processes during their execution - automatic delegation of tasks taking into account the load of agents and their competences.	Improve accuracy
	2	Productivity and quality AI	Ensemble of intelligent interoperable agents works with fewer mistakes and is safer. Ensemble of intelligent interoperable agents improves the quality of life of man and society in daily concerns, as well as productivity in high-tech industry and production.	Improve efficiency
AI features	Task(s)	1 .Safe interaction of ensemble of intelligent interoperable agents. 2 .Building high-tech synergies of ensemble of intelligent interoperable agents..		
	Method(s)	Criterion method of utility and preference, multi-level reflection, semantic and behavioral ontology, of technological ensembles of intelligent agents.		
	Hardware	Supercomputer with Strong Artificial Distributed Intelligence		
	Topology	Distributed Modular Interconnect Topology		
	Terms and concepts used	high-tech synergies, intelligent interoperable agents, utility and preference criteria.		
Standardization opportunities/ requirements	Multimodal multisensory format			
Challenges and issues	Information security			
Societal concerns	Description	Security, ethical and legal aspects		
	SDGs to be achieved	Multi-level processing of big data by intelligent neural systems		

Table 2 Data

Description	Strong Artificial Distributed Intelligence Data
Source	Model and technology of Strong Artificial Distributed Intelligence
Type	Strong
Volume (size)	Hi-Tech Labor Market
Velocity (e.g. real time)	Supercomputering Velocity
Variety (multiple datasets)	streams of multiple datasets
Variability (rate of change)	Retraining
Quality	High

Table 3 Process scenario

N.	Scenario name	Scenario description	Triggering event	Pre-condition	Post-condition
1	Training	Train a model (deep neural network) with training data set	Technological process raw data set is ready	Formatting of data	Management of safety
2	Evaluation	Expansion of the trained model	Development of technological thinking and behaviour	Cognitive thinking patterns and psychological behaviors	Meeting KPI requirements is condition of development
3	Execution	Model and Technology Tooling	Interaction	Activization of Model	Completion of interaction
4	Retraining	Retrain a model with training data set	Certain period of time has passed since the last training/ retraining	Additional data and knowledge	Combining Data and Knowledge

Table 4 Training

Step No.	Event	Name of process/Activity	Primary actor	Description of process/activity	Requirement
1	Sample raw data set is ready	Specification and classification	Manufacturer	Transform sample raw data	Distributed AI Software
2	Completion of Step 1	Creating Set of Experimental Data	Manufacturer	Development of set of experimental data through job modelling	Software of modelling
3	Completion of Step 2	Model training	AI solution provider	Train a model (deep neural network) with experimental data set created by Step 2	Big SD

Table 5 Evaluation

Step No.	Event	Name of process/Activity	Primary actor	Description of process/activity	Requirement
1	Completion of training/retraining	Research	Manufacturer	Train model (deep neural network) with experimental data set created	Big SD
2	Completion of Step 1	Identification	AI solution provider	Based on data, detect execution using a deep neural network trained in learning scenario	Big SD
3	Completion of Step 2	Evaluation	Manufacturer	Comparison of phase 2 results with human performance	Efficiency and quality
Input of evaluation		Productivity			
Output of evaluation		Efficiency and quality			

Table 6 Execution

Step No.	Event	Name of process/Activity	Primary actor	Description of process/activity	Requirement
1	Comparison of modeling results with human performance	Research	Manufacturer	Development of a set of experimental data through job modelling	Quality
2	Completion of Step 1	Identification	Manufacturer	Based on modified data train model (deep neural network) with experimental data set created	Compatibility
Input of Execution		Modification			
Output of Execution		Compatibility			

Table 7 Retraining

Step No.	Event	Name of process/ Activity	Primary actor	Description of process/activity	Requirement
1	Certain period of time has passed since the last training/retraining	Research	Manufacturer	Additional data and knowledge	Completeness
2	Completion of Step 1	Experimental data set creation	Manufacturer	Combining Data and Knowledge Based on modified data train model (deep neural network) with experimental data set created	Compatibility
3	Completion of Step 2	Model training	AI solution provider	Comparison of phase 2 results with human performance	Efficiency and quality
Specification of retraining data		Retraining data set has to include recent data			

5 Conclusion

Currently, multi -modal generative technologies of artificial intelligence continue to effectively transform various industries [9–11]. Generative artificial intelligence is constantly improving and approaching in cognitive abilities to natural intelligence [12]. Natural intelligence builds vital activity on the basis of a productive system of rational and moral meanings approved by the practice of life. Productive meanings are active memory elements. Based on them, thinking is built in actualized situations and circumstances. Thinking is carried out on the basis of meanings of holographic memory, taking into account time and space. When researchers of artificial intelligence will be able to carry out universal standardization of modeling productive semantic thinking of natural intelligence by self -organizing intellectual systems based on rational and moral meanings of their bioinformation holographic memory, then strong artificial intelligence will become an indispensable complement of human natural intelligence [13–17].

Conflicts of interest

The author declares no conflict of interest.

References

[1] Creating a safe AI: specification, reliability, guarantee, 2018. https://habr.com/ru/articles/425387

[2] ISO/IEC JTC 1/SC 42/WG 4 Use cases and applications Convenorship: JISC (Japan). 2019-12-23. https://isotc.iso.org/livelink/livelink/open/jtc1sc42wg4

[3] Bryndin EG. Standardization of artificial intelligence. Standards and Quality. 2020, 12: 22-25.

[4] Bryndin EG. Development of behavioral and professional skills of sensitive cognitive robots as an aspect of safety. /X International Scientific Conference ”IT - STANDARD 2020” - M.: Prospekt Publishing House. 2020: 303-310.

[5] Bryndin E. Standardization of Artificial Intelligence for the Development and Use of Intelligent Systems. Advances in Wireless Communications and Networks. 2020, 6(1): 1. https://doi.org/10.11648/j.awcn.20200601.11

[6] Bryndin EG. Formation of an ethical smart digital environment of industry 4.0. /XI International Scientific Conference “IT - STANDARD 2021” - M.: Prospekt Publishing House. 2022: 6-13.

[7] Bryndin E. Development of Artificial Intelligence of Ensembles of Software and Hardware Agents by Natural Intelligence on the Basis of Self-Organization. Journal of Research in Engineering and Computer Sciences.2023, 1(4): 93-105.

[8] Bryndin E. Development of Artificial Intelligence for Library Activity and Industrial and Social Robotization. Chapter of book “Application of Artificial Intelligence in Library Services”. Springer. 2024.

[9] Bryndin E. Creation of multimodal digital twins with reflexive AGI multilogic and multisensory. Research on Intelligent Manufacturing and Assembly. 2024, 2(1): 85-93. https://doi.org/10.25082/rima.2023.01.005

[10] Bryndin E. Network Training by Generative AI Assistant of Personal Adaptive Ethical Semantic and Active Ontology. International Journal of Intelligent Information Systems. 2025, 14(2): 20-25. https://doi.org/10.11648/j.ijiis.20251402.11

[11] Bryndin E. Intelligent Digital Clinic of Interacting Multimodal AI Assistants. Research in Medical & Engineering Sciences. 2025, 11(4): 1237-1241.

[12] Bryndin E. Formation of reflexive generative A.I. with ethical measures of use. Research on Intelligent Manufacturing and Assembly. 2024, 3(1): 109-117. https://doi.org/10.25082/rima.2024.01.003

[13] Bryndin E. Cognitive Resonant Communication by Internal Speech Through Intelligent Bioinformation Systems. Budapest International Research in Exact Sciences (BirEx) Journal. 2023, 5(4): 223-234.

[14] Bryndin EG. Development of artificial intelligence of ensembles of software and hardware agents to natural intelligence based on self-organization. Yearbook “Greater Eurasia: Development, Security, Cooperation”. 2024, 7(2): 42-49.

[15] Bryndin E. Creation of Multi-purpose Intelligent Multimodal Self-Organizing Safe Robotic Ensembles Agents with AGI and Cognitive Control. COJ Robotics & Artificial Intelligence. 2024, 3(5). https://doi.org/10.31031/cojra.2024.03.000573

[16] Bryndin E. Self-learning AI in Educational Research and Other Fields. Research on Intelligent Manufacturing and Assembly. 2025, 3(1): 129-137. https://doi.org/10.25082/rima.2024.01.005

[17] Bryndin EG. Digital Doubles with Reflexive Consciousness in Reality and Virtual Environment. Greater Eurasia: development, security, cooperation: materials of the VII international scientific and practical conference, Part 2. Moscow: Publishing house UMC. 2025: 380-384.