Journal of Digital Economy Research. 2023. 3. P. 48–84
DOI: 10.24833/14511791-2023-4-48-84

Received: 21.09.2023
Revised: 19.10.2023.
Accepted: 12.11.2023

HIERARCHY OF ETHICAL PRINCIPLES FOR THE USE OF ARTIFICIAL INTELLIGENCE IN MEDICINE AND HEALTHCARE

Valeria N. Sokolchik, Aleksey I. Razuvanov

Abstract. The article researches the problem of ethical support of the application of artificial intelligence (AI) in medicine and healthcare, which is topical for modern science. Despite a significant number of foreign and domestic publications devoted to the topic of AI, the conceptual justification of the ethics of AI application in medicine and healthcare remains poorly developed. Relying on international recommendations and articles, as well as on their own experience of research activities, work in research ethics committees, the results of a pilot survey of health care workers, etc., the authors define and analyze the basic ethical principles of using AI in medicine and health care. The proposed principles are considered in the context of their practical application to protect human and natural rights and interests, which includes preservation of patient confidentiality, prevention of discrimination, protection from AI errors, respect for informed consent, as well as compliance with the norms of "open science", mutual trust of developers and users, etc. The proposed principles are analyzed in the context of their practical application. The application of the proposed principles will orient scientists, AI developers, ethical committees conducting expert review of research, society as a whole to the priorities of humanization of healthcare, respect for human beings and nature, as well as to educate society, create a regulatory framework, ethical recommendations and codes of ethics for the use of AI in medicine and healthcare.

Keywords: AI, artificial intelligence, medicine, healthcare, open science, ethics, manageability, safety, explainability, effectiveness, fairness, trust, ethics committees

For citation: Sokolchik V.N., Razuvanov A.I. (2023). Hierarchy of Ethical Principles for the use of Artificial Intelligence in Medicine and Healthcare. *Journal of Digital Economy Research,* vol. 1, no 4, pp. 48–84. (in English). DOI: 10.24833/14511791-2023-4-48-84

Information about the authors:

Valeria N. Sokolchik, PhD in Philosophy.
Institute of Philosophy of the National Academy of Sciences of Belarus.
ORCID: 0000 0002 4975 4052
vsokolchik@mail.ru

Aleksey I. Razuvanov, PhD in Medicine.
State Institution "Republican Scientific and Practical Center for Medical Expertise and Rehabilitation".
ORCID: 0000-0001-5033-2933
doc-rai@yandex.by

Introduction. The use of AI today is the basis of almost any activity. Especially relevant is the use and development of AI systems for application in the field of medicine and healthcare, because it allows to qualitatively improve and enhance the system of assistance to the population, improve the processes of rehabilitation, medical prevention, expertise and, in general, the organization of the healthcare system. No less significant than the actual development of AI systems for application in medicine and healthcare is the ethical support of the relevant processes of AI creation and application. Since we are talking about humans, their interests, sensitive data, life goals and attitudes, etc., the most balanced and careful attitude to human personality, respect for its autonomy, fairness in the organization of assistance, absence of harm and, of course, trust in AI systems, becomes the cornerstone of AI implementation in medicine and healthcare. Without diminishing the importance of developing legal systems (legislation) regulating the use of AI, it is necessary to emphasize that, first, the most significant and successful legal solutions “get matured” within the framework of ethics, second, if it is impossible to prescribe in a legal document all possible mechanisms and algorithms for the use of AI, we must have an ethical basis on which we can rely in the absence of a legal solution and, third, ethical documents and postulates are extremely important for AI training.

No less relevant is the ethical support for the application of AI in medical science (research and testing). This is particularly important for the new paradigm of “open science”, whose existence is essentially determined by interaction with AI. Open science relies on machine intelligence both in methodology, technologies used, and in analyzing “big data”, processing the results obtained, as well as in disseminating research data through open sources (such as scientific platforms, social networks, etc.) [7]. The use of AI in science, especially those related to the study of human, society, and the biosphere, a priori requires humanistic support of research, assuming the formation of ethical and legal guidelines that define the boundaries and limits of AI use. No modern scientific research, in particular biomedical research, can be carried out without following ethical norms, since it involves the use of one’s personal data, interference in his/her personal life, manipulation with his/her body, broadcasting of his/her opinions, etc. The tools for ethical correction of scientific research are, firstly, ethical literacy and training of the scientist and research team, secondly, research ethical committees (in Belarus, these are independent ethical committees (IEC), which carry out ethical review of scientific projects, thirdly, the society itself, represented by organizations and individuals who influence the formulation of questions, research design, dissemination of research data, etc.).

The purpose of the proposed article, prepared as a result of an interdisciplinary initiative to comprehend AI ethics, is to analyze the ethical basis for the use of AI in medicine and health care, building a hierarchy of relationships between the ethical principles of working with AI, the ethical problems that arise in the application of these principles, and the ways to resolve them relevantly by modern science and practices.

Materials and Methods. It is important to note that although the sphere of medical activity and health care organization as a whole always contains a significant component of ethical and legal regulation, including official documents (both international and national), recommendations, existing ethical standards and structures (ethical committees and commissions), etc., however, innovations related to the use of AI in medicine and health care are not yet sufficiently studied, therefore, there are no clear ethical guidelines yet, and the issues of ethical support of the use of AI (and AI elements) are not yet clear.

The work on the formation of ethics of AI use, in particular, the ethics of scientific research using AI in the modern world is very active. Over the last 3-5 years, a large number of documents have appeared that offer ethical regulations and recommendations for developers and users of AI systems. Among the most significant are the recommendations of the Council on Artificial Intelligence of the Organization for Economic Cooperation and Development; ethical recommendations of the International Coalition of Regulatory Authorities in the field of medicines for clinical medicine and pharmaceuticals; statement on artificial intelligence, robotics and "autonomous" systems of the Council of Europe; recommendations on the ethical use of AI by UNESCO, WHO guidelines on ethics and management of AI use in healthcare, the Code of Ethics of AI of the Russian Federation and others[5; 6; 7; 13; 15; 22; 23] . European project implementing the construction of a platform for European open science is under development, including the creation of ethical standards for research with the use of AI, as well as recommendations for ethical review of research with the use of AI by the Ada Lovelace European Institute of Science [14; 19].

For all the seeming diversity of publications and documents on AI ethics, it is necessary to state that such publications prevail in Western Europe, America and Russia, while in Belarus the issues of research ethics with the use of AI are practically not considered. The methodological problem of studying AI ethics is the lack of reflection on the identification of basic principles of AI ethics, the designation of which would make it possible to create a foundation on the basis of which the mechanisms of ethical reflection and learning of both humans and AI are developing.

In preparing the material, the authors relied on the study of contemporary documents, recommendations, codes, scientific projects and articles, as well as on interviews with AI developers and users in the field of scientific research, a pilot survey of health care professionals, their own experience of scientific and practical activity in medicine and health care, and the experience of expertise of ethical committees of the health care system [8; 11].

The results obtained and their discussion. The global internal problem of ethical support of AI in medicine and healthcare was identified by the authors during interviewing and surveying healthcare professionals: it lies in the respondents' lack of understanding of the terminology related to AI, as well as the essence of the concepts used. Interestingly, the definition of AI caused significant difficulties for the participants of the survey (even with the proposed definitions).

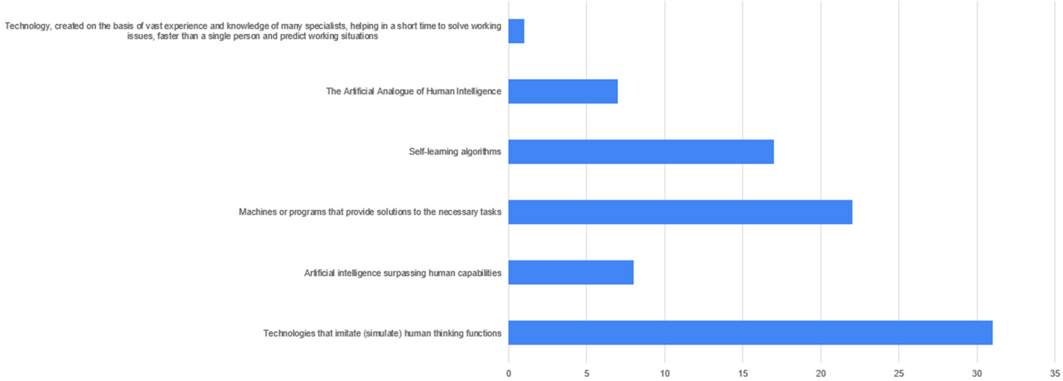

Figure 1. Understanding of AI by survey participants

As authors consider, the discrepancy in the answers and insufficient understanding of the object of consideration is, firstly, due to insufficient literacy and lack of education of medical workers on the development of AI technologies. Secondly, with the fundamentally different level of implementation of relevant technologies in different countries and regions, including in connection with the "gap" of AI technologies implementation in advanced clinics, centers (medical universities) and technologically backward structures.

However, even if we rely on classical "educational" definitions, regardless of our perception of AI - in its "weak" version (as machines and algorithms which solve specific tasks) or in its "strong" version (as an artificial analog of the mind capable of multifunctionality and self-learning) - ethical principles of construction and relationship with AI become a necessary basis for AI technologies, building a framework of (self-) limitation of AI aimed at preserving human and natural existence, at improving the quality of life

Hierarchy of ethical principles of AI use in medicine and healthcare and principles of bioethics. The ethical imperatives underpinning the use of AI in modern medicine and healthcare are the basic principles of modern bioethics. They accumulate the orientation of all actions for the benefit of human and nature ("do good"), the idea of doing no harm to the living ("do no harm"), as well as the recognition of the right of a human to preserve his or her values and determine the boundaries of his or her identity (the principle of autonomy). These principles, regardless of the sphere of their application (practice or science, treatment or prevention, accompanying processes related to human recovery, etc.) remain the main reference points in medicine and public health care, reminding of the main values of the human personality, the supreme goal of medicine and public health care - restoration and preservation of human health, as well as the ethical imperative of the great German philosopher I. Kant (to act in such a way that human is always an aim, but not a means).

In the context of creation and further working with AI to realize basic bioethical imperatives, it is also necessary to think through the principles governing the interaction between AI and humans, as well as their interaction and hierarchy.

The actual set of ethical principles governing the use of AI in medicine and healthcare was proposed by participants in the above-mentioned pilot survey

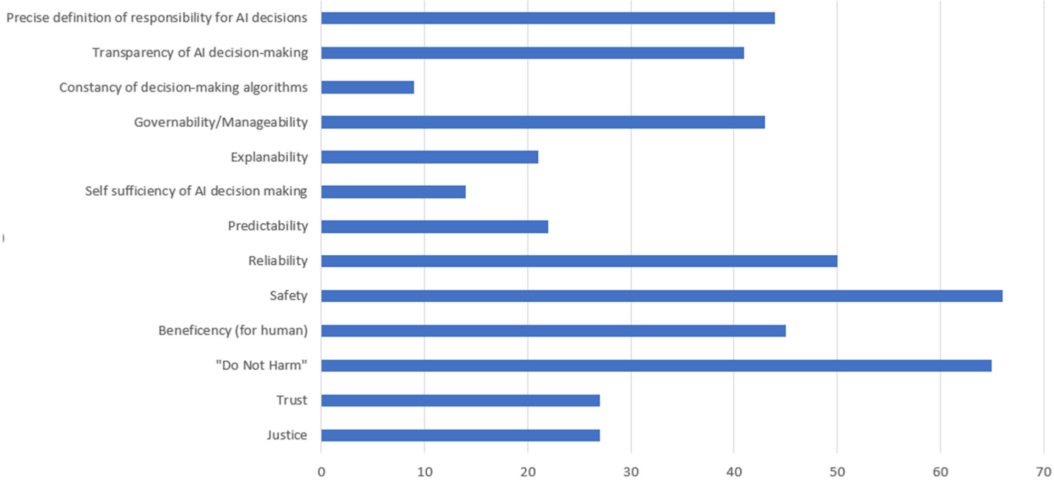

Figure 2: Ethical principles for the use of AI as suggested by respondents in the survey

Objective principles. *The first block* of ethical principles of work with AI - namely, safety (for human, society, nature), controllability, explainability, and efficiency - can be conditionally called "objective". This block defines supra-personal ethical requirements imposed directly on AI systems, ethical and legal attitudes of society, as well as criteria of scientificity and validity of AI use.

Of course, the proposed list of principles can be expanded and supplemented, but the outlined constructs are basic, necessary both for developers of AI systems and for users of AI - researchers, operators, etc., as well as for independent ethical committees (hereinafter - IEC), conducting expert review of scientific research [10]. In order to explain the above principles, it is necessary to dwell on their functionality in a little more detail.

The safety principle implies taking into account several parameters necessary for correct work with AI systems [1]. First of all, safety includes reliability and predictability of systems' actions, including a clear definition of the criteria of AI reliability for humans, society and nature, pre-determination of the system of analysis and correction of errors, and thoughtfulness of the system of protection against risks [17].

When we talk about reliability in the field of research ethics, we do not consider solely the technical reliability of the system. Here we are talking about the development of reliability parameters that determine the safety of humans and nature. In addition to physical safety, which implies the absence of physical harm and discomfort to humans and nature, there is also mental, socio-cultural, etc. safety of humans. It is connected with respect for the human personality's attitudes, careful attitude to his values and ideals, non-disclosure of personal data, respect for the decision made by a person, etc.

It is important for medicine and health care to determine group and personal risks for patients when using AI. While group risks are usually foreseen by the developer (provided that he works together with medical representatives), attention to personal risks remains minimal. Personal risks may arise due to both physical differences of patients and mental, educational, etc. peculiarities (significant both in medicine and in sociology, psychology, pedagogy, etc.), which may be related to inability to use the proposed devices, or uncertainty in sufficient security of their data, etc. The issues of resolving and overcoming personal risks are often solved through informing, instructing, competent explanation and willingness to "go along to get along" in solving the issue. For example, if a patient is unwilling to allow his or her data to be widely used in medical databases, such data should not only be hidden, but also deleted (which in turn requires adequate algorithms and technical solutions). When investigated biological material is used without informing (and authorizing) its owner - this can also be regarded as a direct violation of the security parameter and the necessary criteria of system reliability. A classic example of such a situation is the story of Henrietta Lacks' "immortal" HeLa cells used by the scientific community without her permission (and without informing her family), which subsequently caused serious problems for researchers. Thus, to resolve the situation, together with representatives of the Lax family, it was decided to publish the decoding of the HeLa genome, restricting access to it. If they wish to familiarize themselves with this important data, scientists should apply to the National Institute of Health (USA), where their request will be considered substantively in terms of the objectives of the study and priorities in the dissemination of its results, with the participation of representatives of the Lux family in the committee.

Security also relies on the sustainability of AI systems - i.e., uninterrupted operation of the product taking into account possible external and internal threats, assuming the existence of a self-protection system against such threats and influences. In order to comply with the sustainability parameter, it is necessary not only to provide a list of possible threats, but also to develop a system of appropriate protection (e.g., a system of verification, authentication, user authorization), as well as a system of notification of stakeholders about the occurrence of unplanned influences [9; 27]. To a large extent, sustainability (and, as a consequence, safety) is also determined by the human factor - the actions of users, including doctors and other medical workers, technicians, patients, and other people involved in working with AI systems. In this case, the necessary condition for ensuring sustainability is careful instruction, training of users and constant monitoring of compliance with the instructions received.

Thus, consent to the use of data, availability of mechanisms (algorithms) ensuring restriction of access to data, availability of a multilevel system of user protection, together with protection from AI errors and distortions, protection from unauthorized interference in the system's activity, the possibility of setting and removing restrictions for users, maintenance of data integrity, are mandatory criteria ensuring the safety of AI use. On the basis of these criteria, both the developer and the user, including the physician, researcher, patient, test subject, as well as NECs conducting expert review of research, should assess the adequacy of the use of AI.

The principle of controllability. The principle of safety is inextricably linked to the principle of controllability of AI activity. Manageability is considered here as the possibility to control the operation of AI systems and the existence of a clearly defined and hierarchized system of responsibility for actions and results produced by means of AI. Manageability implies:

- Firstly, the initial definition of the required parameters and settings for the application of the system in a particular area. For example, in medicine, a prerequisite for the application of AI systems is compliance with such parameters as physical and mental safety of the patient, the focus of any actions to improve the quality of his/her life, the requirement to inform the patient, the application of existing ethical and legal norms, environmental friendliness, etc.;
- Secondly, governability requires that AI systems (as well as their developers and users) operate strictly within the legal framework, which implies the adoption at the state, regional (and global) level of special recommendations and legislative norms that include legal and ethical requirements for working with AI;
- Thirdly, the manageability of AI systems is ensured by the creation and timely updating of the Digital Code (hereinafter - the Code), which regulates the legal relations of participants in the digital space. At the same time, the possibility of ethical self-learning of AI systems can be realized by connecting to the digital version of the Code;
- Fourthly, the controllability of AI systems is based on the formation of a system of responsibility. Such a system presupposes hierarchy and distribution of responsibility between all participants who work with AI - from the creator to a particular user (database owners, developers, consultants, testers, supporting staff, as well as end users all become bearers of such responsibility). The end users here are understood as medical workers themselves, organizers of the health care system and all those associated with it on a professional (volunteer) basis, as well as patients, participants in research and trials, and even members of their families, who are responsible for the accuracy of compliance with the prescribed technologies, compliance of the actual conditions of AI use with those specified in the prescription (if applicable), and so on. To address these issues, it is necessary at the national level to elaborate a system of relationships between the above actors, clearly define the scope of their responsibilities, basic requirements and duties, providing for the analysis of responsibility for failures and errors at each life cycle of AI with subsequent correction of errors;
- Fifthly, the controllability of AI systems provides for mechanisms for assessing the harm (damage) caused by AI systems with subsequent compensation. This issue should also be regulated by legal norms, including conditions, grounds, determination of the amount and other parameters of compensation [20]. For biomedical research

the problem of compensation is especially relevant. For example, in the 1980s, the Therac-25 radiation therapy machine developed by Atomic Energy of Canada Limited "AECL" due to a failure in computer coding delivered damaging doses of radiation to patients with oncology, which resulted in a fatal outcome. Liability in this case is still being debated, as some hospitals implemented their own system upgrades that may have caused the overdose. Because there are many parties involved in such an AI system (data provider, developer, manufacturer, programmer, developer, user, and the AI system itself), establishing liability in a controversial case is difficult, with many factors to consider [26]. The Therac-25 incident has been called one of the most serious computer errors in history, requiring special attention to the safety of AI for humans, as well as solving the problems of damage assessment and compensation [26].

The principle of explainability is associated with the existence of a fundamental possibility to understand the actions of the system, the transparency of its algorithms, the accuracy of explanation of the essence and results of the process of generating results [24]. The principle of explainability is realized through the criteria of transparency of AI actions, relying on the principles of safety and controllability.

In medicine and healthcare, the principle of explainability is extremely important for users - both medical professionals (healthcare professionals) and patients, as the pattern of a "black box" generating results undermines the credibility of the research and the belief in its significance. If the researcher does not understand the general algorithm of obtaining results through AI, he/she stops managing the process, noticing errors and, consequently, the results obtained become exclusively quantitative, difficult to meaningfully interpret. Errors made by AI are quite difficult to realize, because they are fundamentally different from those which could be made by a human being. This topic, in particular, is discussed in Xiaoxuan Liu's article "Bringing AI to Responsibility", where the author emphasizes that AI errors cannot always be foreseen, corrected, or even understood by humans [18].

A paradoxical example of errors in AI actions during diagnosis is given in a paper by Lauren Ockden-Reiner and her colleagues, who analyzed the productivity [21]. This study identified several "failure modes", i.e., the propensity of AI to make periodic errors under certain conditions. Among the most significant "failures" related to deviations from the set parameters, one case was mentioned where the AI "missed" a severely displaced femoral neck fracture (according to the authors, even a layman would recognize such an image as completely out of the norm). If the doctor had not double-checked the data, a serious error in diagnosis and the next phase of research (treatment) would have been made, and it is not clear who should be held responsible for such an error.

AI errors in database compilation and statistical processing of results can be no less catastrophic. For example, in May 2018, IBM's digital assistant Watson recommended incorrect and health-threatening drugs to cancer patients. The problem was the system's use of incorrect algorithms: instead of processing patient data and synthesizing new treatment ideas on that basis, Watson was found to be using hypothetical data. Watson's suggestions were based on the preferences of a few physicians who provided data for the development of the system, rather than on real conclusions obtained from analyzing a large number of clinical cases [20].

Lack of transparency in explanation undermines trust in the medical professional and the health care system as a whole: on the one hand, by surrounding the result with a halo of mystery, on the other hand, by depriving it of objectivity and scientificity. The inability to explain the process of obtaining a result to the structures that carry out ethical review of research/quality review of medical care usually leads to a denial of the rational value of the result, a denial of its scientificity, and a requirement to double-check the results.

The principle of efficiency prescribes the accuracy and precision of data collected and processed by AI systems. An important efficiency criterion is compliance with such a parameter as "knowledge limit", which limits the operation of AI systems to the conditions and purposes for which the system was designed. I.e., unauthorized change of system operation parameters by users (e.g., change of temperature regime, operating rules, scope of operation, etc.) threatens the emergence of unpredictable errors, distortions and, as a consequence, data failure.

The effectiveness of AI systems also depends on the users' ability to work with the system or with data. Sometimes incompetence leads to the fact that the unambiguity of AI analysis results and predictive capabilities of the system are over-exaggerated, resulting in "the appearance of objectivity and neutrality of the choice by justifying the latter by the 'infallibility' of the AI analysis" [5]. [5].

The efficiency of AI systems is determined not only by protection against errors and distortions, but also by the validity of the analyzed data. For example, Saracci, a famous researcher of AI, considered validity as maximum unification and standardization of approaches and algorithms of data acquisition, constant analysis of possible sources of distortions. Also, validity requires the collection of critical, important data, which is determined by researchers and experts in the field of solving the tasks [24]. Although the issues of data validity are primarily relevant for scientists offering scientific and practical conclusions based on data analysis, it should be remembered that in the absence of validity, comparison and contrast of data will not be at least correct, and the final product will not meet the considerations of safety and protection of patient rights.

Subjective principles. The above "objective" ethical principles of work with AI (safety, explainability, efficiency, controllability), according to the authors, should be supplemented by "subjective" principles of work with AI, which, as already mentioned above, are largely related to the attitudes of society and individuals, determined by the level of education, ethical and legal norms, attitude to scientific research, the degree of prevalence of the use of AI, etc.

The second block of principles, including the principle of justice and the principle of trust, can be conditionally called "subjective", since it is determined to a greater extent by the internal attitudes of an individual (society), associated with the inclusion in the process of interaction with AI and personally colored perception of this process.

Principle of fairness. Fairness as an ethical principle implies the possibility of equal access to the use of AI systems, as well as the possibility of not interacting with AI or replacing AI systems with human actions. For example, many patients (research participants) may not have access to the electronic devices/programs being used, may also lack the knowledge and skills to use different AI technologies in the absence of assistance (this is especially true for the elderly), etc. Consequently, for many research participants and patients, the use of various medical programs, devices, etc. up to and including booking an appointment with a specialist on a website may be difficult and sometimes even impossible. Addressing issues related to equality of access requires the researcher to clearly identify vulnerable groups for whom the access to AI systems may be difficult or impossible for various reasons, and to initially provide for the possibility of replacing AI algorithms with human actions. For example, to provide an opportunity to order booking to a specialized doctor by phone (without using the Internet), or, for example, instead of using special equipment (devices) at home, to provide the right to perform the necessary actions in a polyclinic under the supervision of specialized specialists, etc.

Undoubtedly, with the development of AI technologies, the education of society and human ability to use AI technologies are developing in parallel, but we should not forget that today quite a large percentage of people are not capable, do not have the physical ability or for value reasons do not want to work with AI systems, but it is not humane to deprive them of quality care. It is important to understand that provision of equal access of patients to treatment and other means of health care presupposes the availability of special knowledge and skills of health care workers, including the ability to competently explain and teach how to operate the necessary AI systems, inform about the use of devices, etc.

The problem of inequality of access to AI systems, which, in fact, becomes the basis for AI discrimination, is complemented by another implicit problem, which can be defined as "latent" discrimination. The latter is related to the cognitive bias of developers which is transferred to the AI system. While "explicit" discrimination is quite easy to detect by a researcher or medical professional, "latent" discrimination is much more difficult to deal with. For this purpose, the physician, researcher, as well as the expert, consultant will have to evaluate the system from the point of view of non-discrimination, eliminate the possibility of inequality of opportunities and ensure that the algorithms work in accordance with the ethical principle of fairness. For example, to evaluate and change (if possible) the parameters of the system which gives inequality of access, results on the basis of race, nationality, gender, political views, religious beliefs, age, social and economic status or privacy information. According to M. Pizzi and his colleagues, AI algorithms can "reinforce existing inequalities between people or their groups, as well as exacerbate the disadvantage of certain vulnerable demographic categories" [5]. [5].

The principle of fairness thus places serious demands on the developers of AI systems, requiring not only to analyze the system for the presence or absence of discrimination, but also to take measures to verify the data set used for machine learning. An important task becomes the creation and application of methods and software solutions that initially prevent the occurrence of discrimination, as well as the adaptation of algorithms in accordance with national priorities and preferences, mentalities, and the specifics of national (regional) health care systems. As noted in the report on the role of AI in humanitarian research "...a system developed in Silicon Valley but deployed in a developing country may not be able to take into account the unique political and cultural characteristics of that country. The developer may not be aware that in the country N certain stigmatized populations are underrepresented or invisible in the datasets, and therefore will not correct the training model for this bias " [5]. Thus, the main ethical recommendations for implementing the principle of equity are that developers should pay attention to the specifics of the model they will be working with, including its sociocultural characteristics. Constant consultations of developers with specialists of the professional sphere for which the program (algorithm) is developed are also necessary. It is also important to develop competent and understandable instructions for working with a particular AI system.

On the part of health care professionals, to ensure the principle of equity, it is necessary to improve their skills (level of knowledge) in the context of working with AI, to follow the instructions precisely and to carry out mandatory pre-testing of the system's operation, identifying risks with regard to possible discrimination of patients, followed by the implementation of appropriate corrections (if necessary).

The principle of trust. This is the most complex ethical principle due to its "subjectivity" and relativity - the principle of trust, on which the relationship between humans and AI systems is based. This principle in relationship between humans means confidence in a person, his integrity, sincerity of intentions and, as a consequence, formation of relations based on such confidence. However, in the situation with AI, both humans and "inanimate" AI participate in the interaction system, therefore, along with "confidence", the subjective feeling of reliability and acceptance is also significant for the formation of trust.

Understanding trust involves at least three "slices" of the relationship:

- trust of consumers of information products to developers, assuming constant and direct solution of problems arising with them;
- trust of developers to consumers, understood as confidence in the responsibility, literacy, thoroughness of consumers in relation to AI, strict adherence to prepared instructions;
- trust of consumers in AI itself, expressing most often a subjective emotional attitude of a human to AI.

All these three “layers” of trust are based on the solution of several basic problems. First of all, it is the problem of user data security, which becomes maximally relevant in the context of medicine and healthcare [14].

Fundamental to resolving the issue is a clear definition of the concepts of “identifying data” (data through which a specific human can be identified) and “non-identifying”, “reversibly anonymizable”, pseudo-anonymizable, personal, etc., data used in modern scientific discourse.

The next step is the formation of the possibility to use personal data (with the permission of its owner) and personal information in a limited way, i.e., the construction of AI systems taking into account the restriction and regulation of access to data. For example, in medicine, including biomedical research, the issues of access to patient data become extremely sensitive: this is associated, in particular, with the problem of possible access to medical history, screening results, etc. information about patients by specialists who are not directly involved in research or medical care (employees of information services, developers, personnel servicing AI systems, etc.), as well as by specialized doctors, medical personnel not related to treatment and examination of patients, and also by specialists who are not involved in the development of AI systems.

This problem is especially complicated in connection with the use of big data, which is a huge amount of information about each patient or subject (not only medical information), accumulated by AI systems through social networks, indicators of various devices, and even through telephone conversations. Of course, analyzing such complex data is absolutely invaluable for medicine, but the personal data risk of becoming open for unauthorized persons is quite high. For example, when undergoing genetic testing, a patient receives very important information, possibly even determining his or her future (hereditary diseases, the possibility of having children, predispositions to diseases, etc.), but the leakage of such information can literally ruin his or her life. Thus, the issue of regulating/restricting access to data is complemented by the requirement of data protection through the development of algorithms that are as free from external and internal risks as possible.

The second basic issue related to trust is informed consent. Informed consent of the patient (hereinafter - IC) is a prerequisite for any action in medicine and health care, IC for complex interventions is understood as a formal document where, based on the information provided about the intervention to be performed, possible risks and potential benefits, the patient gives his or her permission (consent) to participate in [3; 4]. Unfortunately, even carefully designed consent does not always meet the ethical requirements of working with AI systems. For example, the patient is not always informed or asked his/her will about further storage and use of the biological material taken from him/her, screening and diagnostic data, which is necessary for the creation of databases, not always informed about the use of AI systems (e.g., during surgery), not to mention the usual questionnaire with the use of AI, etc. The IS should also include explanations about the possible use of mobile applications (if applicable), explaining that these devices are not intended for use in medical diagnosis, monitoring, treatment, rehabilitation, do not meet the information security requirements under the principles of "designed security" and "designed privacy" [2].

And, of course, the decision of the individual that he or she expresses in the IP regarding his or her data and biological materials must be steadfastly respected. Various types of informed consent, including those that allow the appointment of a 'third party' to subsequently decide on the use of data and materials, or to restrict use and destroy data, have already been sufficiently developed within the research ethics, and the authors do not consider it necessary to dwell on this issue in detail [16].

Another problem related to AI trust, according to the authors, is the regularly encountered "fear" or rejection of AI by humans. The solution to this problem lies in educating and enlightening society, creating reliable AI algorithms, and building human confidence that the final decision always rests with the individual. We are talking about individual's unconditional right to dispose of his or her data, to make decisions based on the information provided and personal values. In the context of biomedical research, it is important to emphasize the right and responsibility of human beings to make important decisions about life and health. Examples of such "human" decisions in biomedicine are those which determine the quality of life, health and existence in society - e.g., diagnosis, palliative care decisions, allocation of limited life support resources, etc. AI can replace human labor or analytical activities, but in most cases, AI cannot replace humans "when making decisions on particularly sensitive issues or on problems without addressing which significant negative consequences may occur" [24] [24].

Addressing human distrust of AI and sometimes aggression towards AI is also provided through:

- transparent analysis of errors made by AI and their subsequent elaboration (which is closely related to the implementation of ethical principles of safety, explainability, manageability);
- providing for "human" alternatives to AI systems (e.g., alternatives to the use of "bots");
- assessment of the professionalism of program developers within the framework of access to the creation of a socially significant product, coupled with interdisciplinary discussion of such programs;
- mandatory incorporation of ethical imperatives and algorithms into AI programs;
- engaging independent experts for highly qualified evaluation of AI systems.

Summarizing the explanation of the above-mentioned principles of AI use in medicine and healthcare, the authors visualize their ideas in the figure, suggesting a hierarchy of relevant principles: from the principles of bioethics - to objective principles of AI use - and, further, to its subjective principles.

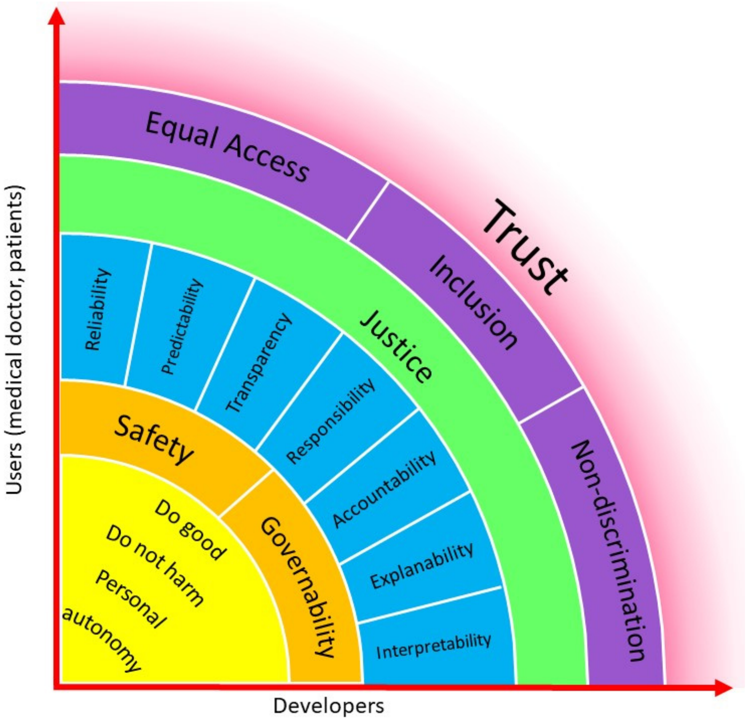

Figure 3: Hierarchy of ethical principles of AI support

Conclusions. Summarizing the results of the conducted research, the authors believe that explication and analysis of the main ethical principles of AI support in medicine and healthcare in the context of safe, mutually beneficial and conflict-free interaction between AI, humans, society and nature is the first step to comprehend the new challenges of modern healthcare, which allows us to create a basis for solving problems related to both the development and training of AI and the organization of modern medical care.

Along with the suggestions addressed to AI developers and users, which have already been made in the article, we consider it necessary to emphasize the following important points:

- the solution of problematic issues related to the use (improvement) of AI should be interdisciplinary practically at all stages of the AI life cycle, because interdisciplinarity allows to combine the efforts of specialists in the field of AI creation, specialists in the professional sphere for which the AI system is developed, as well as specialists in the field of ethics, law, psychology, etc., who ensure the adaptation of AI to work in society and human adaptation to work with AI;
- the creation of ethical codes for AI is now necessary for all participants of the AI-human interaction process. Despite the fact that in recent years many organizations of the world, both local and supranational, have been developing such codes, the

problem remains open due to the fact that the created codes are not based on common principles, are often created for the values and needs of a particular organization or project and have no legal force [16]

- a necessary step for the development of AI is the creation of a legal framework based on ethical principles of work with AI, which defines the responsibilities, rights and obligations of the parties - participants in the process of development and use of AI;
- public education about AI, continuous training on AI in the health care system is a necessary foundation for the development of modern medical science and practice [11];
- an important role in the ethical support of the use of AI in medicine and healthcare should be assigned to ethical committees, commissions (or other ethical structures). In the system of biomedicine, these are, first of all, NECs which provide ethical expertise of research, assess compliance with human rights and rights of nature, compliance with ethical and legal principles of research [10]. Unfortunately, the existing NECs in the field of medicine are not always ready to qualitatively conduct ethical support and expert review of research organized with the use of AI, due to insufficient knowledge about AI, undeveloped ethical (and legal) norms of research with the use of AI, lack of appropriate recommendations, etc. In other professional spheres related to human, social, and natural research (sociology, pedagogy, psychology, history, ecology, etc.) in our country there are actually no ethical structures for research expertise. To date, there are no guidelines or legal basis for them, and accordingly the actual practice of ethics committees outside medicine is extremely rare. Establishment and operation of ethics committees is a modern requirement for the development of science in our country.

Summarizing the ideas developed in the article, it should be noted that the topic of AI research is still open for study, and of particular importance today are the issues of AI ethics. In the context of medicine and healthcare, working with AI requires ethical structures which are capable to advise a healthcare professional, conduct expert examination of the obtained materials, promote the dissemination of scientific knowledge in society, etc. The development of a paradigm for such structures and the study of best practices is a topic which requires further research and development.

References

1. Aseeva I. A. AI and big data: ethical problems of practical use (analytical review) // Social and Humanities. Domestic and foreign literature. Ser. 8, Naukovedenie: Abstract journal. - 2022. - №. 2. - P. 89-98.

2. Bryzgalina, E.V., Alasania K.Y., Varkhotov T.A., Gavrilenko S.M., Ryzhov A.L., Shkomova E.M. Biobanking: socio-humanitarian aspects. - Moscow: Moscow University Publishing House. - 2018. - 232 p.

3. Goloborodko N.V., Sokolchik V.N., Aleksandrov A.A. Recommendations on obtaining informed consent for participation in scientific research: textbook - Minsk: BelMAPO. - 2020. - 36 p.

4. Voluntary informed consent (2022) / Scientific Editor A.G. Chuchalin, E.G. Grebenshchikova. - Moscow: Veche. - 2022 - 288 p.

5. AI in humanitarian action: human rights and ethics // icrc.org URL: https://international-review.icrc.org/sites/default/files/reviews-pdf/2021-12/IRRC_913_pp18734_Article_by_Pizzi_Romanoff_Engelhardt_RU.pdf (accessed 11.11.2023).

6. Code of Ethics in the field of AI // AI Alliance Russia URL: https://ethics.a-ai.ru/ (accessed 11.11.2023).

7. Preliminary Draft UNESCO Recommendation on Open Science // unesco.org URL: https://unesdoc.unesco.org/ark:/48223/pf0000374837_rus (accessed 11.11.2023).

8. Razuvanov A.I., Sokolchik V.N. Ethical challenges of the application of AI in medicine and medical research // Voprosy organizatsii i informatizatsii proizvodstva. - 2023. - №2 (115). - P. 43 - 95.

9. Romanova I.N., Naumov O.V. Application of AI in healthcare // Sustainable development: research, innovation, transformation : Proceedings of the XVIII International Congress with elements of scientific school for young scientists. In 2 volumes, Moscow, April 08-09, 2022 / Volume 1. - Moscow: Moscow University named after S.Y. Witte. S.Y. Witte. - 2022. - P. 460 - 468.

10. Sokolchik V.N. The role of ethics committees in ensuring human rights in biomedical research and testing in the Republic of Belarus // Proceedings of the Belarusian State Technical University. - ser. 6 №1. - 2021. - C. 146 - 150.

11. Sokolchik V.N.. Open science as a new paradigm of scientific research: problems and prospects (on the example of biomedical research) // Proceedings of BSTU. Series 6: History, Philosophy. - 2023. - №1 (269). - P. 163 - 169.

12. Shnurenko I. AI on the verge of a nervous breakdown // Expert. - 2109. - №1-3 (1103). - P. 39-42.

13. AI in medicine regulation News 16/08/2021 // European Medicines Agency URL: https://www.ema.europa.eu/en/news/artificial-intelligence-medicine-regulation (accessed 11.11.2023).

14. Blueprint for an AI Bill of Rights // The White House URL: https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (accessed 11.11.2023).

15. Ethics and governance of AI for health // WHO URL: https://www.who.int/publications/i/item/9789240029200 (accessed 11.11.2023).

16. Fjeld J, et al. Consensus in Ethical and Rights-based Approaches to Principles for AI // Berkman Klein Center Research Publication. - 2020 - №. 2020-1. P. 37 - 48.

17. Hagendorff, T. The ethics of AI ethics: An evaluation of guidelines // Minds and Machines. - 2020. - V. 30. - №. 1. - P. 99-120. URL:https://doi.org/10.48550/arXiv.1903.03425

18. Liu X., et al. The medical algorithmic audit. // The Lancet Digital Health. - 2022. – V. 4, Issue 5, May 2022. P. 394-397 URL:https://doi.org/10.1016/S2589-7500(22)00003-6

19. Looking before we leap. Expanding ethical review processes for AI and data science research // Nuffield Foundation URL: https://www.adalovelaceinstitute.org/report/looking-before-we-leap/ (accessed 11.11.2023).

20. Naik N., et al. Legal and Ethical Consideration in AI in Healthcare: Who Takes Responsibility? // Front Surg. - 2022. - № 9, 266 p.

21. Oakden-Rayner L., et al. Validation and algorithmic audit of a deep learning system for the detection of proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study // The Lancet Digital Health. - 2022. - Volume 4, Issue 5. - p. 351 - 358. URL: https://doi: 10.1016/S2589-7500(22)00004-8

22. OECD Legal Instruments // OECD URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL (accessed 11.11.2023).

23. Recommendation on the Ethics of AI // UNESCO.ORG URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455.locale=en (accessed 11.11.2023).

24. Reddy S. Explainability and AI in medicine //The Lancet Digital Health. - 2022. - T. 4. - №. 4. P. 214 - 215. https://doi: 10.1016/S2589-7500(22)00029-2)14. № 33(3). - P. 245-257. https://doi: 10.1007/s10654-018-0385-9.

25. Statement on AI, robotics and 'autonomous' systems // europa.eu URL: https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1 (accessed 11.11.2023).

26. The Worst Computer Bugs in History: Race conditions in Therac-25 // BugSnag URL: https://www.bugsnag.com/blog/bug-day-race-condition-therac-25/ (accessed 11.11.2023).

27. Vincent-Lancrin, S., van der Vlies, R. Trustworthy AI (AI) in education // Promises and challenges. - 2020. - P. 117 - 125. https://doi: /10.1787/19939019.